

Categorical sensitivity to color differences

Christoph Witzel

Department of Psychology, University of Giessen,
Giessen, Germany



Karl R. Gegenfurtner

Department of Psychology, University of Giessen,
Giessen, Germany



Categorical perception provides a potential link between color perception and the linguistic categories that correspond to the basic color terms. We examined whether the sensory information of the second-stage chromatic mechanisms is further processed so that sensitivity for color differences yields categorical perception. In this case, sensitivity for color differences should be higher across than within category boundaries. We measured discrimination thresholds (JNDs) and color categories around an isoluminant hue circle in Derrington-Krauskopf-Lennie (DKL) color space at three levels of lightness. At isoluminant lightness, the global pattern of JNDs coarsely followed an ellipse. Deviations from the ellipse coincided with the orange-pink and the blue-green category borders, but these minima were also aligned with the second-stage cone-opponent mechanisms. No evidence for categorical perception of color was found for any other category borders. At lower lightness, categories changed substantially, but JNDs did not change accordingly. Our results point to a loose relationship between color categorization and discrimination. However, the coincidence of some boundaries with JND minima is not a general property of color categorical boundaries. Hence, our basic ability to discriminate colors cannot fully explain why we use the particular set of categories to communicate about colors. Moreover, these findings seriously challenge the idea that color naming forms the basis for the categorical perception of colors. With respect to previous studies that concentrated on the green-blue boundary, our results highlight the importance of controlling perceptual distances and examining the full set of categories when investigating category effects on color perception.

lightness, and saturation. Indeed, we are able to differentiate a continuum of more than two million discernible colors (Pointer & Attridge, 1998; Linhares, Pinto, & Nascimento, 2008). On the other hand, when communicating about colors we do not refer to metric evaluations of hue, lightness, and saturation. Instead we use color terms that refer to more or less discrete color categories, such as blue, purple, or pink. In this way, we collapse the three dimensions of hue, lightness, and saturation onto one categorical level, where the high number of discernible colors is reduced to a handful of basic color categories. Here we test whether color categorization is related to our basic ability to discriminate colors.

Background

This research question has become relevant for at least two research traditions. First, the link between low-level color perception and high-level color appearance constitutes a major question in color science.

Color research has achieved detailed knowledge about the low-level mechanisms of color perception. Two first stages of color processing are distinguished (e.g., De Valois & De Valois, 1993). At the first stage, cones are activated by the wavelength combinations of light that enter the eyes. There are three types of cones: short-, middle-, and long-wavelength cones, or S-, M-, and L-cones (e.g., Schnapf, Kraft, & Baylor, 1987; Stockman & Sharpe, 2000; for review, see Gegenfurtner & Kiper, 2003, pp. 182–184).

At the second stage, the activations of the cones are combined in the retinal ganglion cells to produce a three-dimensional (3-D) color-opponent space (Lee, Martin, & Valberg, 1988; De Valois, Abramov, & Jacobs, 1966; Derrington, Krauskopf, & Lennie, 1984; Krauskopf, Williams, & Heeley, 1982). One of these dimensions corresponds to luminance. It is determined by the sum of the activation of M- and L-cones: $L + M$;

Introduction

The present study tackles the explanatory gap between color perception and color naming. On the one hand, we perceive color continuously in terms of hue,

Citation: Witzel, C., & Gegenfurtner, K. R. (2013). Categorical sensitivity to color differences. *Journal of Vision*, 13(7):1, 1–33, <http://www.journalofvision.org/content/13/7/1>, doi:10.1167/13.7.1.

the contribution of S-cones to luminance is negligible. The second dimension represents the differential activation of M- and L-cones: $L - M$. The third dimension results from the activation of the S-cones in contrast to the activation of the M- and L-cones together: $S - (L + M)$. These dimensions are produced by six unidirectional physiological mechanisms, the so-called second-stage mechanisms (e.g., Gegenfurtner & Kiper, 2003, pp. 184–188).

These three dimensions represent color processing starting from the retinal ganglion cells (Dacey, 2000; G. D. Field et al., 2010) via the lateral geniculate nucleus (LGN; De Valois, Abramov, & Jacobs, 1966), up to the primary visual cortex (V1; Chatterjee & Callaway, 2003; Lennie, Krauskopf, & Sclar, 1990). The actual neuronal implementation of the second-stage mechanisms is more complicated than a simple contrast between cone types (G. D. Field et al., 2010) and is not yet completely understood (e.g., Tailby, Solomon, & Lennie, 2008). Nevertheless, the most recent findings (G. D. Field et al., 2010) still support the idea that the second-stage mechanisms result in the 3-D, color-opponent space described above (e.g., Demb & Brainard, 2010). In this way, the second-stage mechanisms provide the basis for our ability to see and distinguish colors (Gegenfurtner & Kiper, 2003, pp. 184–188; Stockman & Brainard, 2010; Boynton & Kambe, 1980).

However, these three dimensions do not suffice to represent color perception at higher levels of processing (e.g., Giesel, Hansen, & Gegenfurtner, 2009; Hansen & Gegenfurtner, 2013; for review, see Gegenfurtner, 2003). In fact, the high-level processes that ultimately lead to color appearance are barely understood (for review, see Krauskopf, 1999; Stockman & Brainard, 2010). As a consequence, color research has turned to investigating the computations on the color signal that are performed by high-level cortical mechanisms (e.g., Brouwer & Heeger, 2009; Parkes, Marsman, Oxley, Goulermas, & Wuerger, 2009; for review, see Eskew, 2009; Gegenfurtner & Kiper, 2003; Gegenfurtner, 2003; Valberg, 2001). From this perspective, the present study is part of the endeavor to establish a link between low-level color perception and high-level color appearance.

Second, the present research question is central to research on the role of language in perception and cognition in general. The idea that language affects perception and cognition has been formulated in the Sapir-Whorf hypothesis, which is named after the authors who inspired this thread of research (Sapir, 1921; Whorf, 1964; Gentner & Goldin-Meadow, 2003). In this field, color naming has gained an exemplary status due to the obvious discrepancy between basic color perception and communication. From this perspective, the influence of language on a basic

perceptual property such as color would illustrate the impact of language upon perception and cognition in general (R. W. Brown & Lenneberg, 1954). In color naming research, the main ideas of the Sapir-Whorf Hypothesis may be rephrased as two questions: Firstly, which factors shape the color categories used in communication, and secondly, to what extent is color perception and cognition influenced by perceptual learning due to linguistic constraints (for review, see Kay & Regier, 2006; Roberson & Hanley, 2007)? However, the origin of color categories and their relationship to continuous color perception is still unknown.

Objective

Evidence for categorical perception bridges the gap between color perception and color categorization. According to the idea of categorical perception, the presence of a category border boosts the perception of difference between different colors (e.g., Bornstein & Korda, 1984). For example, two colors on either side of the blue-purple category should appear to be more different or less similar than two comparable colors within the purple category. Hence, there should be a category-specific effect on how we perceptually distinguish colors, namely a *category effect*. In this way, evidence for category effects provides a link between continuous color perception and color communication through categories.

The quest for category effects has been the major approach towards understanding the relationship between color perception and categorization. Category effects within the color domain have been shown in memory (e.g., Bornstein, 1976; Boynton, Fargo, Olson, & Smallman, 1989; K. Uchikawa & Shinoda, 1996; Pilling, Wiggert, Özgen, & Davies, 2003; Roberson, Davies, & Davidoff, 2000; experiment 3b in Roberson, Davidoff, Davies, & Shapiro, 2005), color-term learning (Laws, Davies, & Andrews, 1995; experiment 5 in Roberson et al., 2000; experiment 3c in Roberson et al., 2005), reaction times (Bornstein & Korda, 1984; Witthoft et al., 2003; Daoutis, Pilling, & Davies, 2006; Gilbert, Regier, Kay, & Ivry, 2006; Drivonikou et al., 2007; Winawer et al., 2007; Yokoi & Uchikawa, 2005; Yokoi, Nishimori, & Saida, 2008; Holmes, Franklin, Clifford, & Davies, 2009), infants' attention to differences within a habituation paradigm (Bornstein, Kessen, & Weiskopf, 1976; Franklin & Davies, 2004; Franklin, Pilling, & Davies, 2005), and for the subjective appearance of color differences (experiment 1 in Kay & Kempton, 1984; experiment 4 in Roberson et al., 2000; experiment 3a in Roberson et al., 2005).

Depending on which kinds of category effects are found at which level of color processing, the results

reveal a relationship between color categories and the respective performance, e.g., memorization, at a certain level of processing, e.g., subjective color appearance. The aforementioned category effects concern performances that all require our very basic ability to discriminate colors. The question therefore arises whether these category effects are actually inherent to our sensitivity to color differences (Lindsey et al., 2010; A. M. Brown, Lindsey, & Guckes, 2011).

This idea is supported by Regier, Kay, and Khetarpal (2007) observation that similarity in CIELAB space was relatively high within and relatively low between the categories. Moreover, Özgen and Davies (2002) trained participants with a new categorical distinction within green and blue, and found category effects on color discrimination at the learned category border.

Other studies did not find category effects for color discrimination. There is evidence that discrimination thresholds in CIELUV space do not decrease at the boundary between green and blue (Roberson, Hanley, & Pak, 2009). It has also been shown in CIELAB that there is no effect of the blue-green category border on perceptual grouping (Pinto, Kay, & Webster, 2010). Bachy, Dias, Alleysson, and Bonnardel (2012) measured discrimination thresholds for sinusoidally modulated spectral power distributions and did not find category effects on discrimination, either. Danilova and Mollon (2012, 2010) found that sensitivity for color differences increases towards pure yellow and blue in MacLeod-Boynton space (MacLeod & Boynton, 1979). This result is contrary to the idea of enhanced sensitivity at the category boundary.

Those contradictory findings might be due to the particularities of color sampling and possible failures of the perceptual measures to represent color differences adequately. First, apart from Bachy et al. (2012) and Danilova and Mollon (2012, 2010) all those studies used the Munsell System, CIELUV, or CIELAB to measure perceptual similarity. These measures are only very coarse approximations of color appearance (Hunt & Pointer, 2011; Fairchild, 1998, pp. 219ff; Kuehni & Schwarz, 2008, pp. 167ff). Moreover, the maximally saturated Munsell chips used by Regier, Kay, and Khetarpal (2007) are not equally distant to one another due to the variations in Munsell Chroma (Fairchild, 1998; Munsell Color Services, 2007b, 2007a). At the same time, the absence of category effects in CIELUV and CIELAB in other studies (Roberson et al., 2009; Pinto et al., 2010) may be due to the fact that these spaces coarsely equate perceptual differences in terms of discriminability. Furthermore, all these color spaces are mathematical approximations that involve nonlinear transformations of low-level color information (e.g., Hunt, 1980). For these reasons, results in these color spaces do not inform us about the relationship between categories and low-level color processing.

To address these issues we examined color sensitivity at the level of the second-stage mechanisms to relate color categories to known neuronal processes of color perception. There is evidence that color sensitivity is shaped by second-stage mechanisms and by later cortical mechanisms of color discrimination (e.g., Giesel et al., 2009; Krauskopf & Gegenfurtner, 1992). However, the relationship between these factors and categorization is yet unknown. It is possible that such determinants of sensitivity combine with category effects, rather than producing them. As a result, the local effects of single categories might be hidden by effects of other determinants. Therefore, it is necessary to inspect the overall pattern of sensitivity across colors and account for modulations of JNDs that are not due to category effects. In order to draw firm conclusions about category effects on color discrimination, it is necessary to investigate a larger sample of adjacent categories. For this reason, we examined categories and sensitivity for all hues at fixed levels of lightness and saturation.

Lightness plays an important role in color categorization. At different lightness levels, the same hues may belong to different categories. For example, some red hues become pink when increasing lightness, and brown hues become orange or yellow. This also implies that the location of the categories across hues changes with lightness (e.g., figure 8 in Olkkonen, Witzel, Hansen, & Gegenfurtner, 2010). Genuine category effects in color sensitivity would imply that changes of color categories across lightness correspond to changes in color sensitivity. Consequently, we tested for category effects at different lightness levels, for which the same hues belong to different sets of categories.

We focused on colors that are isoluminant with their background because such colors allow the sensitivity for hue differences to be assessed independent of the modulation of color perception through luminance differences. In the natural environment we mainly use color categories to describe object surfaces. These object surfaces are either typically darker or lighter than the adapting average color, thus we chose two further conditions with a black and a light background. This enabled the actual luminance of the stimuli to be unchanged between conditions, whereas the percept differed.

Previous studies have shown that stimulus presentation on a black background yields the set of categories for light colors independent of the actual luminance of the test colors (Shinoda, Uchikawa, & Ikeda, 1993; H. Uchikawa, Uchikawa, & Boynton, 1989). At the same time, this mode of presentation affects color discrimination (e.g., supplementary material in Witzel & Gegenfurtner, 2011). If categories are firmly grounded in color sensitivity, category effects should also appear when sensitivity changes. In this case, changes in

sensitivity across the lightness conditions should coincide with corresponding changes in categorization.

Another crucial issue that has been neglected in previous studies is the impact of individual differences. If there are individual differences in color categorization and color sensitivity, it makes a fundamental difference whether we analyze individual or aggregated data. Individual differences may be related in two opposite ways to category effects. On the one hand, category effects may be specific to individual patterns of color sensitivity and categorization. This implies that they occur at different locations in color space, depending on individual patterns of categorization and discrimination. Aggregating data here across individuals would hide categorical patterns behind statistical noise.

On the other hand, since categories are used for color communication they may be adapted to patterns in color sensitivity across individuals. In particular, it is possible that color categories are arranged to optimize communication between individuals with differences in color sensitivity (Jameson & Komarova, 2009b, 2009a; Komarova & Jameson, 2008; Komarova, Jameson, & Narens, 2007). As a result, category effects should follow collective rather than individual patterns of sensitivity and categorization. In this case, aggregated data would be more informative. For these reasons, we carefully account for both individual differences and the general patterns in color categorization and color sensitivity in the present study.

In sum, the question about the relationship between color categorization and the sensitivity for color differences is still open. To answer this question, we examined whether there are category effects on color discrimination at the level of the second-stage mechanisms. Such category effects would show that sensory information at these early stages of color perception is further processed so that color discrimination is categorical. In this case, categories would be inherent to color sensitivity, and color perception itself would be genuinely categorical (see also conference abstract, Witzel, Hansen, & Gegenfurtner, 2008a).

To answer the question about categorical sensitivity we compared color categorization to color discrimination. The main results of this comparison will be presented and discussed in the third and fourth sections of this paper, respectively. However, before that, it is useful to evaluate discrimination and categorization separately, in view of their differences across lightness levels and individuals.

Measurements

We modeled the second-stage mechanisms through the Derrington-Krauskopf-Lennie (DKL) color space

(Derrington, Krauskopf, & Lennie, 1984; Krauskopf et al., 1982). We sampled colors of approximately equal saturation along an isoluminant hue circle in DKL space.

We determined the sensitivity as well as the category membership for all hues along this circle at three levels of lightness. First, we assessed color sensitivity through a classical discrimination task that measures discrimination thresholds as *Just-Noticeable Differences* (JNDs). A JND is the minimal difference between two colors that an observer is just able to perceive.

Second, we determined the color categories that correspond to the *basic color terms*. Basic color terms are linguistically defined as the elementary color terms to be used for communicating colors in a particular language (Berlin & Kay, 1969; Crawford, 1982). In English for example, there are eleven basic color terms, namely white, black, red, yellow, green, blue, orange, purple, gray, brown, and pink. Basic color terms can also be empirically distinguished from nonbasic color terms by their comparatively high consistency within subjects, their high consensus between subjects, and their low reaction times in a naming task (Boynton & Olson, 1990; Sturges & Whitfield, 1997; Guest & Van Laar, 2000).

General methods

Color discrimination and categorization were measured on the same setup, and with the same participants. Hence, the following methods apply to both kinds of measurements.

Participants (general)

Eight women (F1–F8) and two men (CW and M2) with an average age of 22 years ($SD = 3.3$ years) participated. Among them, one was one of the authors (CW) and another one was a research assistant of our lab (F8). The eight naive subjects were students at the Justus-Liebig-University in Gießen, Germany, who were paid for participation. All participants were native speakers of German only. Color deficiencies were excluded by means of the Ishihara plates (Ishihara, 2004). Due to the extensive measurements, one participant (F5) dropped out before the end of the experiments, resulting in slightly fewer measurements of discrimination thresholds and color naming for this participant as specified below.

Apparatus (general)

The monitor used to display stimuli was an Iiyama MA203DT CRT monitor (iiyama Deutschland GmbH, Ilm, Germany) driven by a NVIDIA graphics card

(NVIDIA Corporation, Santa Clara, CA) with a color resolution of 8 bits per channel. Spatial resolution was set to 1152×864 pixels and the refresh rate to 75 Hz. For calibration, the spectra of the monitor primaries were measured with a Photo Research PR650 spectrometer (Photo Research, Inc., Chatsworth, CA). The resulting Judd-revised CIE chromaticity coordinates and luminance in cd/m^2 (Judd, 1951) for the monitor primaries were: $R = (0.616, 0.347, 13.8)$, $G = (0.287, 0.606, 41.2)$, and $B = (0.154, 0.078, 5.1)$. For gamma correction, primary intensities were measured with a UDT Instruments model 370 optometer with a model 265 photometric filter (Gamma Scientific [UDT Instruments], San Diego, CA). Based on these measurements look-up tables were created to convert between linear and gamma-distributed RGB values. Experiments were written in MatLab (The MathWorks Inc., 2007) with the Psychophysics toolbox extensions (Pelli, 1997; Brainard, 1997).

Stimuli (general)

We measured JNDs and color categories for colors on a maximally saturated, isoluminant hue circle in DKL space. We sampled stimuli so that hue differences between adjacent test colors were approximately the size of the minimal JND.

DKL space: Figure 1 illustrates the isoluminant hue circle in DKL space. Apart from the scaling of the axes, the isoluminant plane of DKL space is basically the same as the MacLeod-Boynton chromaticity diagram (MacLeod & Boynton, 1979). The three cardinal axes of DKL space represent colors that excite differentially and independently of each other the three perceptual dimensions that correspond to the six second-stage mechanisms (Krauskopf & Gegenfurtner, 1992, pp. 2166–2167; Hansen & Gegenfurtner, 2013). The luminance axis in this space represents colors that activate the luminance mechanisms ($L + M$), while keeping the other two dimensions constant. This is possible since S-cone activation is assumed not to contribute to luminance. For this reason, the luminance axis in DKL space corresponds to the activation of all three cones ($L + M + S$) relative to the white point. The second axis represents the contrast between M- and L-cones ($L - M$), while keeping S-cone activation constant. The third axis defines colors that only excite the $S - (L + M)$ mechanisms. In order to keep the excitation of the ($L + M$) mechanism constant, only S-cone activation varies along this third axis, and hence it is a *tritan* axis.

Rendering: In order to render colors defined in DKL space on the screen, DKL values were linearly transformed into linear RGB values based on the cone fundamentals of Smith and Pokorny (1975; see also Wyszecki & Stiles, 1982, p. 615; Cao, Pokorny, & Smith, 2005). These linear RGB values were gamma-

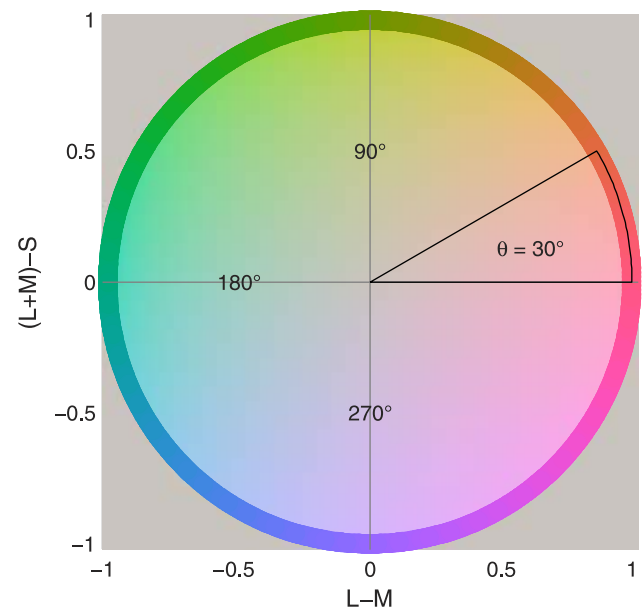


Figure 1. Isoluminant circle in DKL space. Axes are labeled according to the mechanism they activate. The x-axis represents the contrast between L- and M-cones ($L - M$). The y-axis is the tritan axis. It corresponds to the variation in S-cone excitation (high excitation = low y-values). For isoluminant colors, this axis represents the contrast between ($L + M$) and S since ($L + M$) is constant. The origin of this space refers to the adaptation color (i.e., the background). The axes are scaled so that an absolute value of 1 corresponds to the radius of a circle that is tangential to the limits of the monitor gamut. Hence, a radius of 1 defines the most saturated colors at equal radius that are available for a given monitor. We sampled stimuli along this circle. Their hues were defined as the azimuth (θ) in degree along the hue circle. The angle between the black lines illustrates an example azimuth of 30° . The other angles shown in the graphic report the azimuths of the axes.

corrected by means of the look-up tables described above (“Apparatus” section). For details on color rendering in DKL space, see Brainard (1996).

Hue circle: The hue circle for the stimulus sampling was centered around the origin of DKL space, which corresponds to the adaptation point. Hue is specified by the azimuth. Luminance is determined relative to the maximum luminance of the monitor, with 1 being the maximum of the monitor luminance, 0 half of the maximum, and -1 completely black. Finally, increasing the distance of color coordinates from the origin in DKL space leads to increasing the saturation of the color. The relationship between radius and saturation depends on the scaling of the axes. For our study, the two axes were scaled relative to the monitor gamut. A value of 1 and -1 refers to the maximum value along the respective axis that may be displayed on the monitor. To obtain the maximum saturation possible, the hue circle had a radius of 1. As a result, it lies at a

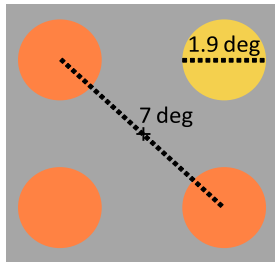


Figure 2. Stimulus display. Four colored disks were presented to the participants in the center of an achromatic computer screen. Three discs show the test, one disc the comparison color. Distances of the dotted lines are indicated in visual angle.

tangent to the lines that correspond to the limits of the monitor gamut.

Stimulus display: Figure 2 illustrates the stimulus display. Stimulus colors were presented as four disks of 1.9° visual angle arranged symmetrically around the center of the computer screen with a diagonal (i.e., maximal) distance of 7° visual angle. Stimulus colors were surrounded by an achromatic background. The colored stimuli were not superimposed upon the background, but were embedded within it. Hence, the chromaticity and luminance of the stimuli did not vary as the background was varied (note that in some contexts, this kind of background might also be called “surround” instead of “background”).

Lightness: The luminance of the stimuli was set to 30.0 cd/m^2 , half of the monitor’s maximum luminance. The luminance of eleven randomly colored stimuli varied between 27.4 and 28.8 cd/m^2 when measured with the spectrometer from the location of the observer’s eye.

However, the lightness of the stimulus colors does not depend on the absolute luminance, but on the luminance difference between stimuli and the luminance to which the observers adapted. All displays (instructions, stimuli, and blank screens) were shown on the uniform, achromatic background of the computer screen, and there was no surround beyond the screen (cf. Figure 2). In this way, we guaranteed that participants adapted to the achromatic background of the screen.

In the main condition, the background was isoluminant to the stimuli. The Judd-corrected chromaticity coordinates of the background were $x = 0.3129$, $y = 0.3484$, $Y = 27.9 \text{ cd/m}^2$ when measured on the screen. In the other lightness conditions, we wanted to obtain the same gamut as for the isoluminant colors. To achieve this we kept the stimulus colors unchanged and changed the luminance of the background. To obtain dark colors we doubled the luminance of the background to the maximum of monitor (59.4 cd/m^2). To produce light colors, we showed the stimulus colors on a black background (0.17 cd/m^2).

Procedure (general)

In one session, participants completed the color discrimination task, followed by (in a few cases also preceded by) either a color naming or a color adjustment task. Each session began with an oral overview of the task given by the experimenter. Detailed standardized instructions were then presented on the computer screen. The time for reading through the instructions guaranteed that participants adapted to the background of the screen. One round of the color discrimination task lasted on average 41 min; the other tasks lasted less than 15 min. Hence, one session took about 1 hr.

Data analysis (general)

If not mentioned otherwise, t tests involved paired samples and were two-tailed. The alpha-error was set to 0.05 ; $p < 0.1$ will be called “marginally significant,” $p < 0.01$ “highly significant.” Univariate Repeated Measures Analysis of Variance (RMAOV) were calculated with the functions of Trujillo-Ortiz, Hernandez-Walls, and Trujillo-Perez (2004) and SPSS (IBM Corp., released 2011). We corrected degrees of freedom by the Greenhouse-Geisser ϵ in order to account for violations of the sphericity assumption (A. Field, 2005, pp. 444–480). We used Fisher’s z -transform to calculate averages and t tests for correlation coefficients (Fisher, 1921, 1915). Reported average correlation coefficients were converted back from z to r .

Discrimination

As explained above it is crucial to examine interindividual differences in color discrimination for the investigation of category effects. Differences in the sensitivity maxima of the L- and M-cones and variations in the second-stage mechanisms are possible sources of sensitivity differences across trichromatic observers (e.g., figure 1 in J. Neitz & Neitz, 2011; M. Neitz, Neitz, & Jacobs, 1991). In fact, it has been shown that discrimination performance along the DKL-axes may vary across trichromatic observers (Webster, Miyahara, Malkoc, & Raker, 2000).

At the same time, color categorization exhibits systematic regularities across different speakers so as to allow for functional communication within a language. If color categorization is directly related to color sensitivity, JNDs should also exhibit a stable pattern across different observers. According to previous studies the presentation of the stimulus colors on the black background should yield different JNDs than the other two lightness conditions (cf. Introduction). For these reasons, we examined differences in JNDs across individuals and lightness levels.

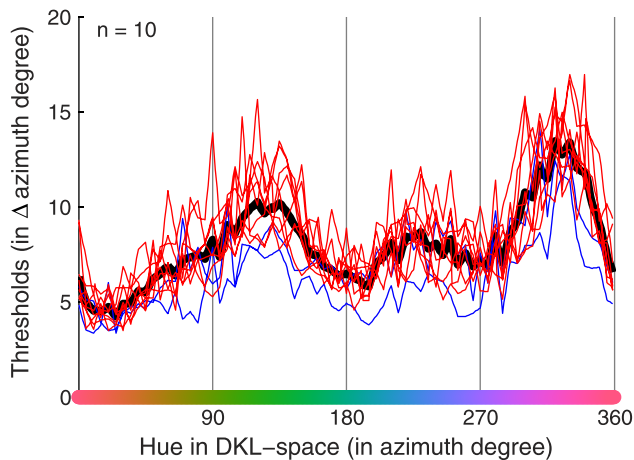


Figure 3. Individual variation of JNDs at isoluminance. The JNDs for each individual are shown as thin colored lines. The x-axis corresponds to the variation in hue as defined by the azimuth in degree. For illustration, the isoluminant color circle of Figure 1 is depicted along the x-axis. The y-axis shows the azimuth differences that correspond to the respective JNDs. Individual JNDs were averaged across the four staircases for each hue at isoluminance (for other lightness levels see Supplementary Figure S4). The eight red curves belong to the female (F1–F8), the two blue ones to the male participants (CW & M2). The black line in the background shows the average across participants. *Note that beyond individual differences and measurement noise, the curves share a common profile with higher JNDs for greenish, bluish, and pinkish hues.*

Method (discrimination)

Participants: All ten participants took part in the measurements of discrimination thresholds with the isoluminant background. For four of the ten participants (F1–F3 and CW), discrimination thresholds were also measured with dark stimulus colors, and for two (F1 and CW) with light colors.

Stimuli: Test colors were sampled at a distance of 5° azimuth around the hue circle starting from 0° azimuth (\sim reddish pink, cf. Figure 1), i.e., 0° , 5° , 10° , [...], 345° , 350° , 355° . This sampling resulted in a set of 72 test colors. As may be verified in the Results (discrimination) section (e.g., Figure 3), 5° azimuth is close to the lowest JNDs. As a result, color sampling covered the complete range of discriminable colors.

Procedure: In order to measure JNDs, we used the spatial 4-Alternative Forced-Choice (4AFC) discrimination task of Krauskopf and Gegenfurtner (1992). In each trial, the observers were shown four colored disks. One of these colors differed in chromaticity from the three other disks. The observer had to indicate which disk was different by pressing one of four keys that corresponded to the four positions of the disks. The 4AFC technique has been shown to be comparatively efficient for determining discrimination thresholds (Jäkel & Wichmann, 2006).

Each trial began with the presentation of a fixation point for 1000 ms. Then the disks were presented for 500 ms, followed by the presentation of the fixation point on the blank background. The trial ended when the participant pressed a key, be it during the 500 ms stimulus presentation or during the blank screen after the stimulus presentation. Each response was followed by feedback, indicating whether the answer was correct or not. This was done by changing the fixation point from black to gray for 500 ms. If the answer was correct, the gray of the fixation point became lighter than the background; otherwise it became darker than the background. The timing of the pre-stimulus and stimulus presentation was chosen so that afterimages were minimized without influencing the discrimination task.

Among the four stimuli of a stimulus display, the three disks with the same color were shown in the *test color*; the one disk with the different color was shown in the *comparison color*. The difference between comparison and test color was adapted through a 3-up-1-down staircase that ended after seven reversal points. The change of these differences were weighed by a factor of 0.259, which corresponds to the Weber fraction of color discrimination according to previous studies. Three-up-one-down-staircases provide discrimination thresholds at which correct responses are given with a probability of 0.79 (Levitt, 1971, pp. 469–474). For 4AFCs, this corresponds to a probability of 0.72 of seeing the respective color difference.

For each test color we measured the discrimination threshold in each of the two directions of hue change through azimuth-increasing and azimuth-decreasing staircases, yielding clockwise and counter-clockwise thresholds, respectively. One azimuth-increasing and one azimuth-decreasing staircase of a test color were measured together and interleaved in one block. In one session, participants accomplished twelve such blocks. The test colors for one session were spaced by 30° azimuth ($360^\circ/30^\circ = 12$). Across six different sessions, the set of test colors was shifted in 5° steps (0° , 5° , 10° , 15° , 20° , and 25°) to measure all $6 \times 12 = 72$ test colors. These measurements were repeated once, resulting in a total of 12 sessions per participant. Participants completed these sessions on different days and the chronological order of the test color sets (i.e., of the 5° shifts) was randomized. Overall, these measurements consisted of 288 staircases. These comprised four staircases (two directions \times two measurements) for each of the 72 test colors. All participants completed all 12 sessions except for one (F5), who participated in 11 sessions (i.e., one repetition is lacking). Four participants were also measured at the dark lightness level (white background). Of these, three completed four sessions and one completed three sessions for this lightness level. The two participants who also did the

color discrimination at the higher lightness level (black background) took part in only two sessions at this lightness level.

Results (discrimination)

When calculating JNDs, we discarded the first reversal point of each staircase to avoid artifacts. Then we averaged over the six (three in each azimuth direction) remaining reversal points of each staircase (Levitt, 1971). We report JNDs as differences in azimuth. Calculating JNDs as Euclidean distances barely changed the JND profile (only 0.1% of normalized JNDs across hues; see also Supplementary Figure S4).

The raw JNDs for each participant may be found in Supplementary Figure S1 of the supplementary material. There we also compare the JNDs resulting from each of the two azimuth increasing and azimuth decreasing staircases (“Differences across staircases” in the supplementary materials and Supplementary Figure S2). In sum, the profile of JNDs was very consistent across different staircases. For this reason, we averaged the JNDs across all measurements so as to obtain a single JND for each test color. Note that for dark and light colors, JNDs were not always measured for the same sets of hues for each participant (cf. Method). To prevent spurious results, tests of aggregated JNDs across hues will only include hues that were measured for all participants in the respective group ($n = 36$ for dark colors, $n = 12$ for light).

Interindividual variability: Figure 3 allows comparison of the single JNDs of each participant for isoluminant colors. For results with different lightness levels see Supplementary Figure S4.

At isoluminance, average discrimination thresholds per individual varied between a minimum of 5.9° and 9.2° azimuth, and the grand average was 8.0° . We calculated a two-way RMAOV to test for differences across hues and observers. The repeated measurements consisted of the four JND measurements, and the factors were hue ($n = 72$) and observers ($n = 9$, after excluding the participant who lacked some measurements). There was a main effect of hue ($\epsilon = 0.03$, $F(5.8, 1.8) = 13.5$, $p = 0.008$), and a main effect of observers ($\epsilon = 0.17$, $F(1.4, 4.2) = 12.2$, $p = 0.02$), but no interaction ($\epsilon = 0.005$, $F(2.7, 8.1) = 1.5$, $p = 0.28$).

When shown on a white background (Supplementary Figure S4a), the average discrimination thresholds over the respective four participants was almost the same as for the isoluminant background, namely 8.1° (min = 6.7° , max = 10.9°). A one-way RMAOV across the $n = 36$ hues with the factor “observers” ($n = 4$) showed that the difference between observers was significant ($\epsilon = 0.62$, $F(1.9, 65.2) = 29.7$, $p < 0.001$). When shown on a black background

(Supplementary Figure 4b), the average JNDs for each of the two participants were much higher, namely 11.5° for one and 13.0° for the other participant (average = 12.3°). The difference between observers was not significant in a t test across hues, $t(11) = 0.98$, $p = 0.34$.

Differences across participants notwithstanding, there were also systematic regularities in the JND profile across observers. We used two measures to assess the stability of the JND profile across participants. Firstly, we calculated pairwise correlations across hues for all pairs of participants. At all lightness levels JNDs were positively correlated across individuals. At isoluminance, pair-wise correlations were on average $r = 0.71$, varying between $r = 0.54$ and $r = 0.83$ (all $ps < 0.01$, $n = 72$). For dark colors (Supplementary Figure S4b), the average correlation of JNDs across hues was $r = 0.43$ (min = 0.32 , max = 0.55 , max $p = 0.06$, all other $ps < 0.05$), and for light colors (Supplementary Figure S4c), the JNDs of the two participants were also strongly correlated with $r = 0.72$ ($n = 12$, $p < 0.01$).

Secondly, we calculated the first principal component across hues (i.e., with participants as variables and hues as cases) to assess in how far the overall variance across hues may be represented by a common prototype. Average JNDs and SE were highly correlated across hues (for isoluminant: $r = 0.77$, dark: $r = 0.67$, and light colors: $r = 0.42$; all $ps \leq 0.01$) as well as across participants (isoluminant [$n = 10$]: $r = 0.87$, $p = 0.01$). Hence, we calculated the principal components based on the correlation matrix in order to decouple the principle components across observers from the variation of averages across hues. The first Principal Component represented 74% of the variance across participants for the isoluminant, and 91% for the dark colors.

Aggregated JNDs: Figure 4 presents the aggregated JNDs for each lightness level in a polar plot. Here, JNDs that were only measured for some of the participants for the dark and light colors were included in the aggregation.

For the aggregated JNDs at isoluminance (gray curve), the global minimum was 4.2° and located at 25° . The global maximum was 13.4° and located at 320° . Further local minima (“dips”) occurred at 195° and 275° with JNDs of 5.8° and 6.8° , respectively. Further local maxima (“peaks”) appeared at 120° and 230° with JNDs of 10.2° and 8.7° , respectively. The global minimum of the aggregated JNDs for dark colors (black curve) was 4.5° and occurred for the test azimuth of 15° . A local minimum was located at 190° (average JND of 6.1°). The global maximum appeared at a test azimuth of 300° (average JND of 12.5°). A local maximum occurred for a test azimuth of 150° (average JND of 12.0°). The aggregated JNDs for light colors (white curve) showed minima at 30° (average JND of

5.9°), 150° (average JND of 6.7°), and 270° (average JND of 10.0°). Maxima were located at 105° (18.3°), 255° (17.4°), and 285° (16.0°).

Differences across lightness levels: The aggregated JNDs for dark colors correlated highly with those for the isoluminant colors, namely with $r = 0.77$ ($n = 72$, $p < 0.001$). The aggregated JNDs for light colors correlated significantly with those for isoluminant colors ($n = 36$, $r = 0.35$, $p = 0.04$), but the correlation with the aggregated JNDs for dark colors did not reach significance ($n = 36$, $r = 0.17$, $p = 0.33$). Moreover, t tests across hues showed that the JNDs for light colors were on average larger than those for the isoluminant colors (on average by 4.3°, $t(35) = -7.2$, $p < 0.001$) and those of the dark colors (on average by 4.4°, $t(35) = 6.7$, $p < 0.001$). In contrast, the JNDs for isoluminant and dark colors did not differ significantly (average = 0.11°, $t(71) = 0.65$, $p = 0.52$).

Global patterns: To explore the global pattern of JNDs we fitted an ellipse to the aggregated JNDs. An ellipse reflects the most general pattern of data through the relative scaling of the axes while transitions between data points are completely gradual. We used the direct least-squares fit of Fitzgibbon, Pilu, and Fisher (1999). Figure 5 compares the fitted ellipse to the empirical JNDs at isoluminant lightness. The dark red curve in Figure 5b shows the residuals as the signed differences between the fitted ellipse and the empirical JNDs. These residuals decrease towards 0°, 180°, and 270°. This implies that sensitivity to color differences is higher around the DKL-axes than predicted by the fitted ellipse. However, there is no such minimum at 90°. At lower lightness, the JND pattern resembles still more an ellipse (cf. Supplementary Figure S5a). At the same time, the residuals do not show the pronounced minima at the axes (cf. Supplementary Figure S5b). The JNDs for light colors strongly deviate from the fitted ellipse (cf. Figure 5c), but the residuals do not result in minima around the axes, either (cf. Figure 5d).

Discussion (discrimination)

We observed a general pattern of JNDs across hues. The main effect of hue in the RMAOV indicated that there was a systematic variation of JNDs across the hue circle (cf. Figure 3). The main effect of observers in the RMAOV revealed that the absolute sizes of JNDs differed across observers. However, the profile of JNDs across hues was also very consistent across observers, as shown by the correlations across hues. In fact, the overall profile of JNDs could be represented by over 70% through a common principal component. Finally, correlations indicated that at least the JND profiles for isoluminant and dark colors were very similar.

The global pattern of JNDs in DKL space may be modulated by the second-stage mechanisms (e.g.,

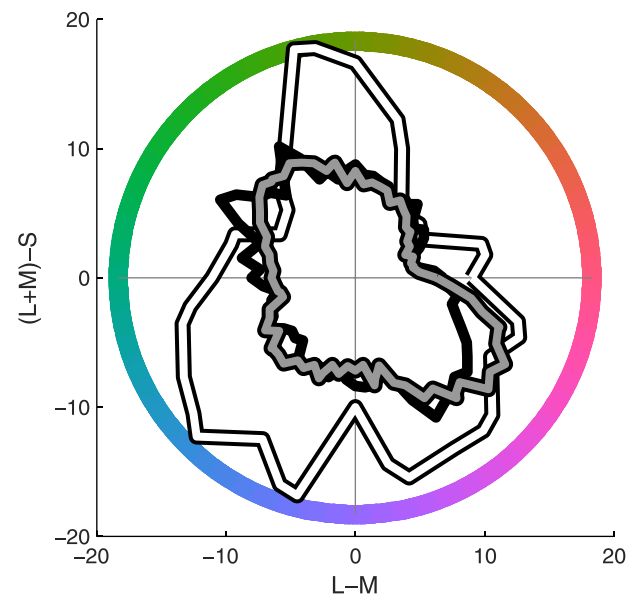


Figure 4. Aggregated JNDs across lightness levels. The direction (azimuth) of the values corresponds to the hues of test colors. These colors are illustrated by the color circle (cf. Figure 1). The size of JNDs is represented by the eccentricity of the values on the curves. The axes correspond to the size of JNDs (as projected on the respective axis). The gray curve shows measurements with isoluminant colors, the black curve those with dark colors (white background), and the white curve the measurements light colors (black background). The measurements with isoluminant and dark colors yielded similar results, but differed from those with light colors.

Krauskopf & Gegenfurtner, 1992) and by cortical mechanisms of color perception (e.g., Giesel et al., 2009). The global pattern for isoluminant and dark colors in the present study roughly followed an elliptical shape (cf. Figure 5a and Supplementary Figure S5a). However, JNDs for isoluminant colors systematically deviated from the ellipse and indicated JND minima around three of the four cardinal directions in the DKL space (0°, 180°, and 270°; cf. Figure 5b). These three peaks of sensitivity are most likely due to the role of the second-stage mechanisms in color discrimination (e.g., Krauskopf & Gegenfurtner, 1992). The second-stage mechanisms do not yield increased sensitivity in the 90° direction, but asymmetries in the 90° and 270° thresholds have been reported before (Krauskopf & Gegenfurtner, 1992; Giesel et al., 2009).

The JND pattern for dark colors almost completely followed the shape of the fitted ellipse (cf. Supplementary Figure S5a), and residuals did not yield minima around the cardinal directions (cf. Supplementary Figure S5a). We speculate that the luminance difference between stimulus colors and background results in interactions between luminance and chromatic channels of the second-stage mecha-

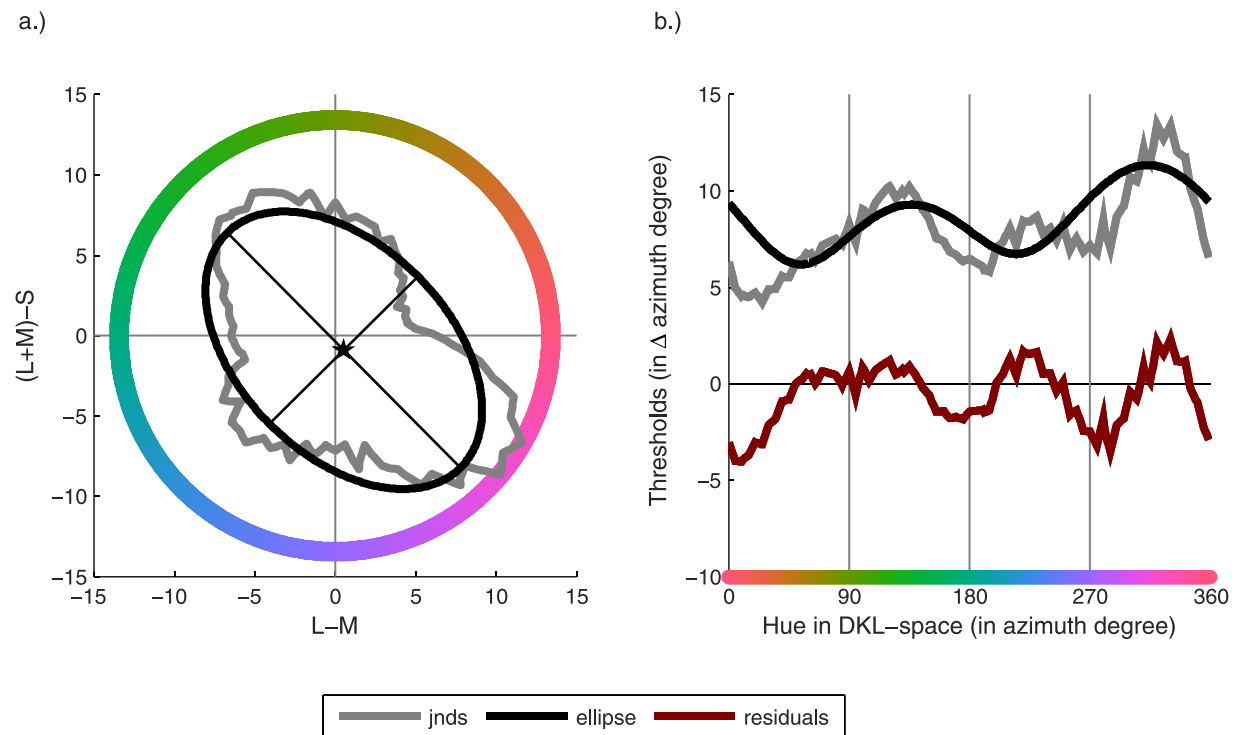


Figure 5. Ellipse fit for JNDs. Based on direct least squares an ellipse is fitted to the aggregated JNDs for isoluminant colors (see Supplementary Figure S5 for other lightness levels). Panel a shows the ellipse (black) and the aggregated JNDs (gray) in the polar representation. The gray curve is the same as in Figure 4. Panel b shows the ellipse (black) and the JNDs (gray) as a function of azimuth in degree along the x-axis (as in the previous figures). The red line represents the residuals—the differences between the ellipse and the empirical JNDs. Differences decrease around 0°, 180°, and 270°, but not at 90°. The relative decrease around the axes indicates an impact of the second-stage mechanisms on color discrimination.

nisms. These interactions might cover or counteract the sensitivity peaks at the chromatic channels.

Moreover, JNDs changed strongly when colors were shown on a black background (cf. Figure 4). JNDs for these light colors were higher overall than for isoluminant and dark colors. Moreover, the profile of these JNDs across hues did not correlate with the profile of the dark colors. It did correlate with the profile of JNDs for the isoluminant colors. However, this correlation was small (12% of total variance), and Figure 4 shows that profiles still differ strongly between these two lightness conditions (see also Figure 9a, c). In fact, the global pattern of JNDs for light colors was much less elliptical than those for isoluminant and dark colors (cf. Supplementary Figure S5).

The unusual distribution of the JNDs for light colors might be due to the contribution of rods under that condition. Under dark adaptation, rods contribute to color perception (e.g., Stabell & Stabell, 1998). The contribution of rods and their interaction with cones affect color sensitivity (e.g., Zele, Kremers, & Feigl, 2012; Knight, Buck, Fowler, & Nguyen, 1998; Stabell & Stabell, 2002). Since rods contribute nonlinearly to color discrimination (e.g., Zele et al., 2012), their

contribution results in a different profile of JNDs in comparison with JNDs for photopic vision (Knight et al., 1998).

The stimulus colors were well in the photopic range. However, between the presentations of the stimulus colors in the lightness condition with the black background, the only light source above 0.2 cd/m² was the tiny white fixation point. The time between stimulus presentations (1–2 s) was certainly not enough to allow for complete dark adaptation, which takes about 8 min. However, it is possible that observers were partially adapted to the dark. This may explain the particular size and profile of JNDs for light colors. In any case, the differences in JNDs between the lightness conditions allow for testing whether category effects are robust to changes in sensitivity.

Finally, we observed that averages and standard deviations of JNDs were correlated across observers and across hues. These correlations may be due to the method since the adaptations of color differences during the staircases were scaled by the Weber fraction (cf. Method), implying higher changes for higher differences during each staircase.

Categorization

Color categories can be characterized by their boundaries and prototypes. A prototype is the most typical exemplar (often also called “best example”) of a color category, such as the typical red. We determined the category boundaries through a color naming procedure, and measured prototypes through a color adjustment task. Since we did not obtain a red category for colors at isoluminance (see Results [categorization]), we verified the absence of red through an alternative naming method.

For the reasons mentioned above, we tested for differences in categorization across individuals and across lightness levels. With regard to the prototypes, an important question was whether the hues of prototypes could be represented at constant luminance along the hue circle. Consequently, we also tested whether the prototypical hues were specific to particular lightness levels.

It is also an important question whether there is a relationship between second-stage mechanisms and color categories. Since sensitivity is related to the second-stage mechanisms, a relationship between categorization and second-stage mechanisms could mediate the one between color sensitivity and categorization. The naming data reported in previous studies did not show a straight-forward relationship of this kind (Malkoc, Kay, & Webster, 2005; Gegenfurtner & Kiper, 2003; Hansen, Walter, & Gegenfurtner, 2007). However, those studies used differently scaled color spaces and different color categories. For this reason, we will reexamine this question with our data. Moreover, the comparison between previous data and ours will allow the general validity of the color categories we obtained to be verified.

Method (categorization)

Color naming: All ten participants took part in the color naming measurements. Stimulus colors were sampled in 3° azimuth steps around the hue circle starting from 0° azimuth, i.e., 0°, 3°, 6°, [...], 354°, 357°, and 360°. This sampling resulted in a set of 120 colors (320/3).

In order to determine color categories, we let participants name the stimulus colors by the eight chromatic basic color terms. This means that black, gray, and white were excluded from the response options because all stimulus colors were highly saturated. Since our participants were German, we used the German basic color terms: Rosa, Rot, Braun, Orange, Gelb, Grün, Blau, and Lila.

The procedure comprised a constant stimuli procedure. In one trial, four colored disks were presented in the same configuration as in the discrimination task (cf.

Figure 2). Here, all disks were colored in the same color. The color was chosen randomly from the stimulus set. In order to assign a basic color term to the color on the screen, the observer had to press one of eight keys, which corresponded to the color terms.

Each trial began with the presentation of a black fixation point for 1000 ms. Then the colored disks were presented until a response was given. At the end, the color name corresponding to the response was displayed for 500 ms to consolidate the association between response keys and color names. After this feedback, a new trial began with the presentation of the fixation point. There were three blocks, one for each level of lightness (isoluminant, white, and black background). In each block, each of the 120 test colors was named once. The order of the blocks was randomized. Except for the one dropout participant (F5), each participant completed this naming task at least five times in separate sessions (F5 did it just once, F1 six times, cf. Figure 6). Consequently there was a data set of five answers for each of the 120 colors with three different background luminance levels per color.

Prototype adjustment: Prototypes were measured for only 9 out of the 10 participants due to one dropout. To determine the prototypes the four disks were presented in a random color, and participants had to adjust their color to the prototype of a given color term by using the cursor keys. There were two conditions. In the first condition, people could only adjust the hue. The background remained isoluminant gray. In the second condition, participants could also adjust the lightness of the colors by changing the background luminance. Since there was no brown at isoluminant lightness, observers adjusted only seven prototypes in the condition with fixed (isoluminant) lightness and all eight prototypes in the condition with adjustable lightness. For details on the adjustment method, see “Prototype adjustments” in the supplementary materials.

Verification of red: In order to verify the absence of red for isoluminant colors, we measured boundaries between pink and red and between red and orange through a 2-Alternative-Forced Choice (2AFC) task. This method forced participants to determine a boundary between pink and red as well as between red and orange, hence establishing a red category. Based on the discrimination data we checked whether these boundaries contained a discriminable amount of color. For details on the method refer to the “Alternative naming task” section of the supplementary material.

Results (categorization)

Naming: Figure 6 shows the raw color naming data for the isoluminant colors. For each hue the mode color name was calculated, and category boundaries (vertical

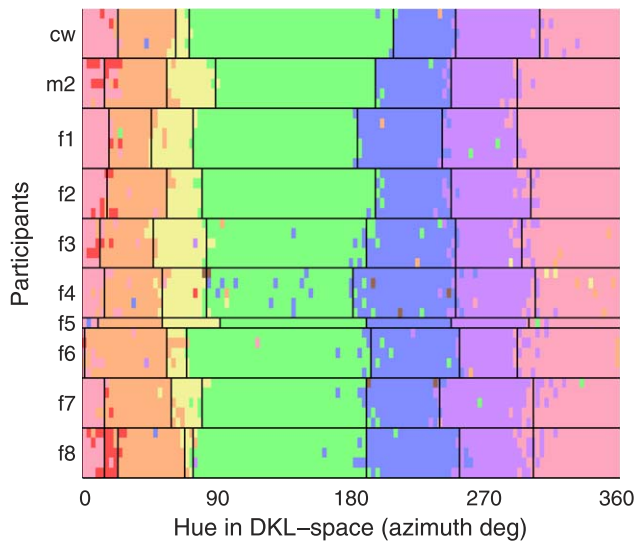


Figure 6. Color naming. The colors in the graphic refer to the color names used for a particular test color in a particular session by a particular observer. The x-axis refers to the variation in hue as in Figure 3. Each column refers to a particular stimulus color at isoluminance. The horizontal black lines separate layers that correspond to the data for each participant. Within each layer, the rows represent repeated measurements of color naming in different experimental sessions. Vertical black lines correspond to the category boundaries calculated through the modes of the single measurements. Supplementary Figure S6 provides the graphic for dark and light. *Variation across was higher than within observers.*

black lines) were determined based on changes in mode names. In equivocal cases (e.g., ties), boundaries were linearly interpolated taking the naming of adjacent hues into account. Analogue graphics for the other lightness levels are provided by Supplementary Figure S6.

Categories for isoluminant colors: Figure 7 highlights the categories resulting from the data in Figure 6 (see Supplementary Figure S7 for other lightness levels). At isoluminance, all participants saw at least the six categories corresponding to the basic color terms pink, orange, yellow, green, blue, purple, and pink. Only one of these participants saw an additional red category. The other participants only saw red at lower lightness (Supplementary Figure S7a). The supplementary 2AFC naming method confirmed the absence of a red category for isoluminant colors. The average distance between the pink-red and the red-orange boundary was 0.8 JNDs. This implies that the category width for red was smaller than one JND, and there are no discriminable colors between the pink-red and the red-orange boundary at isoluminance.

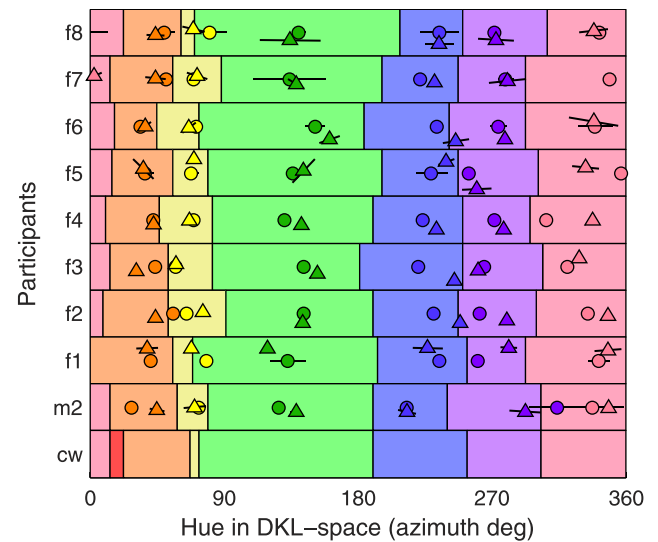


Figure 7. Color categories and their prototypes. The x-axis refers to the variation in hue as in Figure 3. Color categories for each observer are shown as areas in pale colors. The boundaries correspond to the vertical black lines in Figure 6. The symbols refer to the prototypes of the categories as obtained by the prototype adjustments. The saturated colors of the symbols correspond to the respective categories. Disks refer to prototypes obtained from the adjustments at isoluminance, and triangles refer to those with adjustable luminance. The height of these triangles relative to the disks indicates the adjusted luminance. For repeated measurements, SE are shown as thin black lines. Here the results for isoluminant colors are shown, see Supplementary Figure S7 for other lightness levels. *There was (almost) no systematic difference in hue between prototypes at isoluminance (disks) and those obtained with adjustable luminance (triangles).*

Differences across lightness levels: Dark colors yielded a lower consensus in categorization (Figure S7a). The set of categories as well as the boundaries were much more variable across individuals. Only green, blue, and purple existed for all 10 participants. Pink, red, and brown only occurred for 9 of 10 participants. For two participants there was also orange and yellow for one. In sum, only six participants saw the same combination of categories with dark colors. When participants adapted to a black background, they chose the same categories as with the isoluminant background (Figure S7b). However, there were only eight participants with exactly the same set of categories, with 2 of 10 participants naming some colors red.

Interindividual variability: Figure 6 and Supplementary Figure S6 also allow the variability in color naming within and between individuals to be compared. To assess whether color naming differs between individual observers, we tested whether

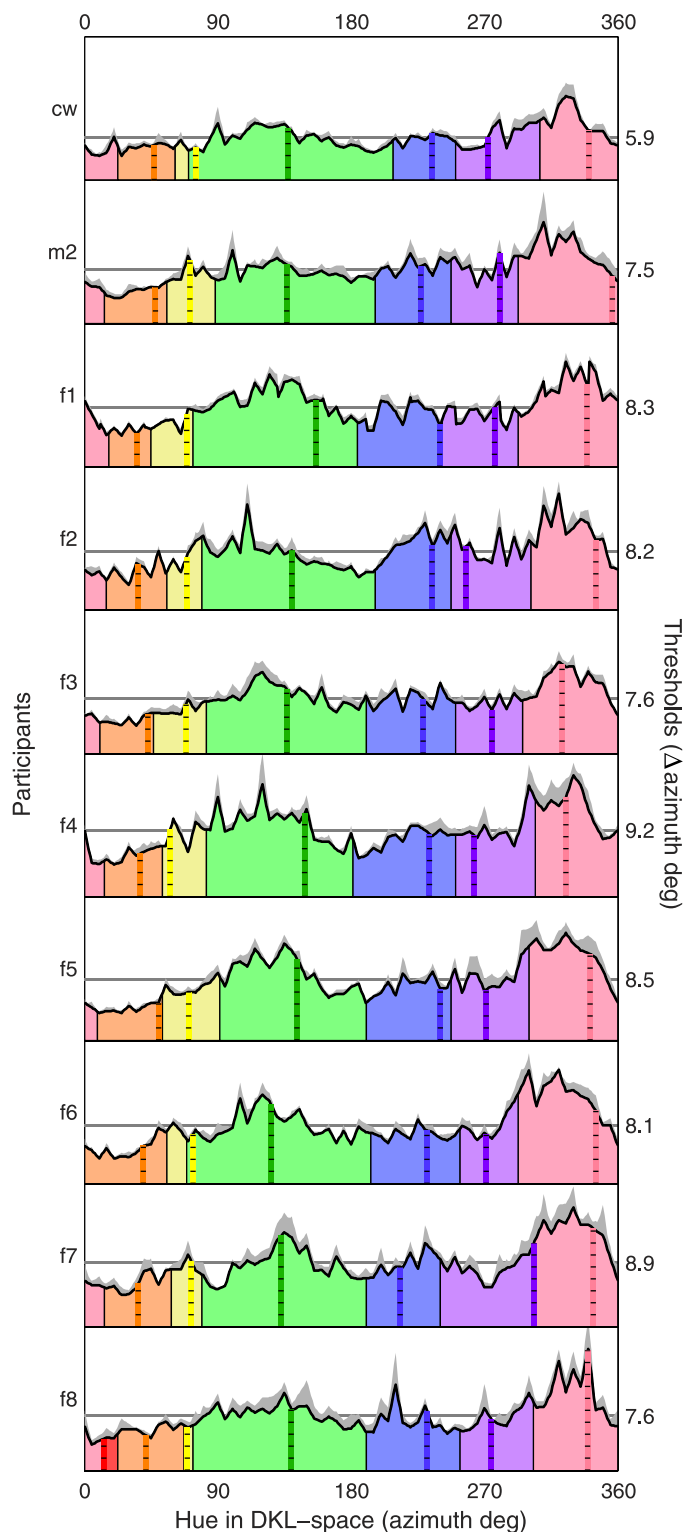


Figure 8. Individual data. The data from categorization (color of areas), prototype adjustments (dotted lines), and discrimination (height of areas) at isoluminance are combined for each individual observer (rows). The x-axis represents variation in hue as in Figure 3. The y-axis on the left side indicates the ID of each observer. The y-axis on the right is split into parts for each individual, and represents JNDs on a scale from 0° to 20°

boundaries and category widths vary more strongly between rather than within individuals. For this purpose, we determined a boundary for each repeated measurement of each individual. We calculated two one-way RMAOVs, one to compare the lower-azimuth boundaries and the other to compare the widths of categories across participants. We excluded the participant with only one color naming measurement (F5). Participants who did not yield a particular category boundary in each of the five measurements were excluded from the analyses of this particular category (at isoluminance only F6 for yellow); we used just the first five measurements for F1 (cf. Figure 6). Table 1 reports the results for isoluminant colors, Supplementary Table S1 those for other lightness levels.

All category boundaries at all lightness levels differed significantly across individual observers (all p s < 0.05; cf. left part of Table 1 and Supplementary Table S1). The only exception were the boundaries between pink and purple when colors were shown on the white and the black background (see first row of Table 1 and Supplementary Table S1, $p = 0.11$ and $p = 0.12$, respectively). Given these differences in boundary locations, it is no surprise that the width of all categories differed significantly across participants, too (all p s < 0.05; cf. right part of Table 1 and Supplementary Table S1). The only exceptions from this were isoluminant and dark pink ($p = 0.06$ and $p = 0.15$), dark red ($p = 0.12$) and dark purple ($p = 0.053$).

Prototypes: The colored discs in Figure 7 and Supplementary Figure S7 represent the average adjustments of prototypes when lightness was fixed at the isoluminant level. Triangles show the adjustments with variable lightness; their relative height indicates the lightness adjustment. These graphics illustrate that the prototype adjustments generally agreed with each other. They also agreed with the categories since there are few prototypes outside the boundaries of the corresponding categories. We provide a detailed evaluation of the adjustments and a comparison of the two different kinds of adjustments in the supplementary material (“Prototype adjustments” section in the supplementary materials).

← azimuth. Each tick on this axis reports the average JND for the respective observer. This average is illustrated by the thick gray lines. The black curve corresponds to the individual JNDs shown in Figure 3. The gray shaded area around the JND curve illustrates the SEM across the four staircases. Categories and prototypes are those from Figure 7. Prototypes were averaged across the two adjustment conditions. Supplementary Figure S7 provides the graphics for the other lightness levels. *The data in this figure was used to test category effects.*

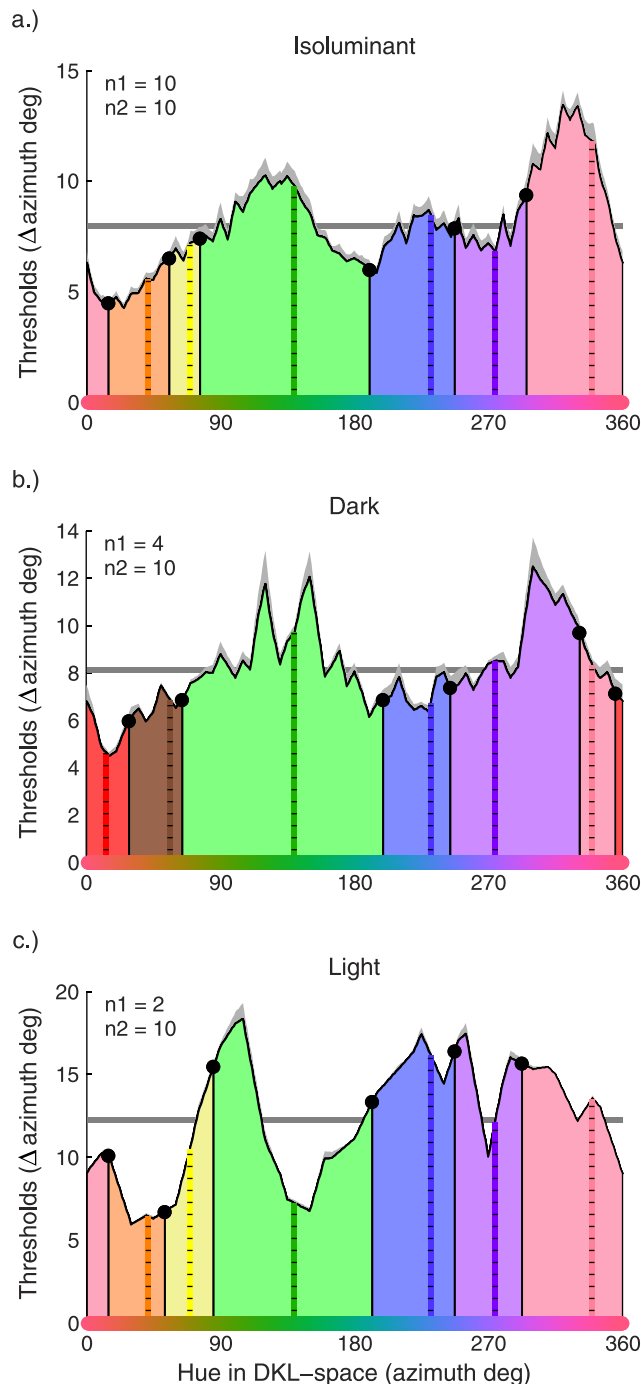


Figure 9. Aggregated data. This graphic shows the average discrimination thresholds (height of colored areas), consensus categories (color of areas), and average prototypes (dotted vertical lines). Analogous to Figure 8, the x-axis shows stimulus hues, the y-axis JNDs, and the thick gray line is the overall average JND. Gray shaded, transparent areas around the JND curve represent *SEM* across observers. Sample sizes n_1 and n_2 correspond to the number of observers for whom JNDs and categories were measured, respectively. The thick black dots indicate the endpoints of the boundary lines (cf. “Predictions and tests” section). Panels a through c shows results with the

Discussion (categorization)

The distribution of categories we obtained for isoluminant colors in DKL space was in line with previous findings. Malkoc et al. (2005) modeled the second-stage mechanisms through MacLeod-Boynton space. For isoluminant colors, they found that the $L - M$ contrast axis was close to prototypical red on one end and to the green-blue boundary on the other. The $S - (L + M)$ axis was close to the yellow-green boundary and the purple prototype, although purple was slightly shifted away from the axis towards more reddish hues (cf. figure 1 in Malkoc et al., 2005, p. 2157).

In our study, prototypical red (8°) and the green-blue boundary (190.5°) were also close to the $(L - M)$ axis of DKL space. Moreover, the $S - (L + M)$ axis was close to the boundary between green and yellow at 76.5° and passed almost exactly through prototypical purple at 272° (see also Figure 9a).

Similar distributions of categories were also observed in other studies that used slightly different color terms to categorize isoluminant colors in DKL space (cf. figure 4 in Gegenfurtner & Kiper, 2003; figure 2 in Hansen et al., 2007, p. 6). The congruence of these results highlights their general validity. Together, they further support Malkoc et al.’s (2005) conclusion that the second-stage mechanisms do neither consistently concur with the prototypes nor delineate the category borders.

However, we also found differences to previous findings. When omitting the pink category, Hansen et al. (2007) found a red category on the isoluminant color circle of DKL space. In contrast, we did not obtain red with isoluminant colors for all participants except one. Our supplementary 2AFC measurement confirmed that there was no red category for our isoluminant colors when using the full set of basic color terms. Even when participants produced red category boundaries in this alternative procedure the difference between the colors at the boundaries was below threshold. We conclude that there is indeed no consistent red category for this set of isoluminant colors. The red category found by Hansen et al. (2007) is most probably due to the fact that participants were forced to call pink colors either red or purple because there was no response option for pink.

As expected, the categorization of dark colors was different from that of isoluminant colors. In particular, dark colors yielded a brown and a red category, but no orange and yellow in most participants.

← isoluminant, white, and black backgrounds, respectively. *JNDs follow a global tendency, which results in categorical patterns for some categories (e.g., green, blue, pink in panel a), but to inverse patterns in others (e.g., purple in panel a).*

Category	Borders					Widths				
	ϵ	df1	df2	F	p	ϵ	df1	df2	F	p
Pink	0.40	3.2	12.7	5.0	*	0.29	2.3	9.2	3.6	°
Orange	0.37	3.0	12.0	14.3	***	0.29	2.3	9.1	10.2	**
Yellow	0.28	2.0	7.8	21.3	***	0.36	2.5	10.0	22.4	***
Green	0.24	1.9	7.7	19.3	***	0.27	2.1	8.5	18.9	**
Blue	0.29	2.3	9.3	7.7	**	0.28	2.2	9.0	8.4	**
Purple	0.26	2.0	8.2	11.6	**	0.40	3.2	12.8	5.6	**

Table 1. Individual differences in categorization. Results of comparing the lower-azimuth boundaries (first part) and widths (second part) of each category across individuals through a RMAOV over five repeated measurements. *Notes:* Number of individuals was $n = 9$, except for yellow ($n = 8$). ϵ refers to the Greenhouse-Geisser correction of sphericity; $df1$ and $df2$ report the degrees of freedom of the numerator and denominator, respectively. Symbols °, *, **, and *** correspond to $p < 0.1$, $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively. See Supplementary Table S1 for other lightness levels.

In contrast, the category sets for light colors were highly similar to those for isoluminant colors. These category sets comprise categories that are typical for light colors, such as pink and yellow, and exclude categories that are typical for dark colors, such as brown. The fact that the condition with the black background yielded the categories that are typical for light colors is in line with previous observations (Shinoda et al., 1993; H. Uchikawa et al., 1989).

The similarity between the categories for isoluminant and light colors indicates that isoluminant colors are also categorized like light colors. This observation is in line with the idea that colors in the natural environment are usually less light than the adapting white point. Since isoluminant colors are as light as the adapting white point, they may appear as particularly light in comparison to the surface colors we categorize in everyday life. This observation also answers the question why red occurs for dark, but not for isoluminant colors. Isoluminant colors are too light to yield a red category.

We also found clear individual differences of category boundaries and widths at all lightness levels. Participants even differed in the set of color terms they used to name the hues. Our results statistically confirmed what previous studies already suggested (e.g., Bornstein & Monroe, 1980, p. 218; Kay & McDaniel, 1978; Witzel, Hansen, & Gegenfurtner, 2008b; Olkkonen et al., 2010, figure 8; see also Hansen et al., 2007, figure 2d; Witzel & Gegenfurtner, 2011).

Some variation of the category sets may be explained by rudimentary categories. For isoluminant and light colors the category sets varied only in the occurrence of red. Category sets of dark colors varied more strongly across participants. In this case, a major source of variation were the categories pink, orange, and yellow.

At the same time, red at higher lightness, and pink, yellow, and orange at lower lightness were also particularly inconsistent across repeated measurements (cf. Figure 6 and Supplementary Figure S6); they yielded many prototype adjustments outside the categories (cf. “Prototype adjustments” section and Figure S7 in the supplementary material). Such inconsistencies of category membership are typical around the category boundaries (cf. figure 8 in Olkkonen et al., 2010). Hence, they indicate that the occurrence of the aforementioned categories varies across participants because only their boundaries reach into those lightness levels.

Apart from these rudimentary categories, we found that the prototype adjustments were generally in line with the categories (cf. Figure 7 and Supplementary Figure S7). Moreover, prototypes tended to vary systematically across individuals (in particular for green and purple). Finally, there was no convincing evidence that one or the other adjustment condition yielded more reliable results (for further details see “Prototype adjustments” section in the supplementary material).

Combination

Figure 8 combines the individual data for color discrimination and categorization. The respective graphics for the other two lightness levels are provided in Supplementary Figure S8.

JNDs of each participant were averaged across the four staircases, and hence the curves in Figure 8 (and Supplementary Figure S8) correspond to those in Figure 3 (Supplementary Figure S3, respectively). For

the colors on the white and on the black background, JNDs were not available for all participants. However, we observed a high interindividual similarity of the profile of JNDs across hues (cf. “Discrimination” section). We therefore used the average JNDs for participants without JND measurements to explore the color categories at these lightness levels. As a result, the JND curve is the same for six and for eight of the 10 participants in Supplementary Figure S8a and b, respectively.

The individual categories shown as colored areas in Figure 8 simply correspond to those in Figure 7. Prototypes are shown as dotted lines in Figure 8. Since we did not obtain convincing evidence that one of the two kinds of prototype adjustments was superior, we used their average hue in order to test category effects. This is the average along the x -axis between the disks and triangles in Figure 7. However, results presented in the following chapters were not different when reanalyzing the data with only one or the other adjustment method. Finally, we used the average prototypes for participant F8.

Results

Figure 9 allows for a first visual comparison between discrimination thresholds and color categories. In this figure, the categories and discrimination thresholds shown in Figure 8 and Supplementary Figure S8 are aggregated across participants (for details about the aggregation see “Aggregated JNDs and consensus categories” section below). Figure 9a suggests that JNDs for isoluminant colors decrease towards the pink-orange and towards the green-blue boundary, and increase around the pink, green, and blue prototypes. However, there are no such signatures at the other four boundaries (orange-yellow, yellow-green, blue-purple, and purple-pink), and the other three prototypes (orange, yellow, and purple). Figure 9b and c shows that there are even less such patterns for dark and light colors, respectively.

However, color categorization and discrimination yielded at the same time systematic regularities and clear differences across individuals (cf. “Measurements” section). The existence of systematic individual variation implies that it makes a fundamental difference whether we investigate category effects for the individual or the aggregated data. Moreover, several different kinds of JND patterns may be predicted based on the general idea of a category effect.

We wanted to make sure that the patterns in Figure 9 are reliable across individuals, and that we did not miss any other signatures of category effects apart from those described above. For this reason, we maximized

the sensitivity for category effects in our analyses. We conceived several tests to cover all possible predictions and we applied these tests to aggregated and individual data, separately. The predictions and the respective tests are presented in the first section. The detailed results of these tests are reported in the second section for aggregated and the third section for individual data.

Predictions and tests

In general, categorical sensitivity should consist of higher discrimination across categories (*similarity expansion*) and lower discrimination within categories (*similarity compression*) as compared to the cone-contrasts produced by the second-stage mechanisms. However, category effects may occur at the category boundaries, the category prototypes, or both. An effect at the boundaries corresponds to the classical idea of categorical perception (e.g., Harnad, 1987; Bornstein & Korda, 1984). An effect at the prototypes is comparable to the idea of a perceptual magnet effect (e.g., Kuhl, 1991). These variants would imply an enhancement of color sensitivity at the category borders (*boundary effect*), a reduction of sensitivity around the prototypes (*prototype effect*), or both. Each of these variants would result in (mainly) three different kinds of *categorical patterns* of JNDs. Hence, we conceived different *categorical perception tests* to detect all kinds of categorical patterns in the JND profile.

Categorical patterns

Categorical patterns are characterized by the relationship of the JNDs within categories (*category JNDs*) to the JNDs at the boundaries (*boundary JNDs*) and to those at the prototypes (*prototype JNDs*). The first row of Figure 10 illustrates the three kinds of categorical patterns. It shows models, which assume that category effects are the only determinants of JNDs. A pure boundary effect implies that JNDs should be lower precisely at the category boundaries. Hence, there should be local minima or “dips” of JNDs at the category borders (Figure 10a). In contrast, a pure prototype effect would result in local maxima or “peaks” of JNDs at the prototypes (Figure 10b). Finally, the combination of both effects would imply that JNDs monotonically decrease from the prototype towards the boundary (Figure 10c). In these ideal models, any category effect would lead to an overall tendency of larger JNDs within categories than at the boundaries.

However, in reality there is a global pattern of JNDs which may combine with possible category effects. In this case, the relationship between the category, boundary, and prototype JNDs also depends on the global pattern. For illustration, the boundary JNDs are shown

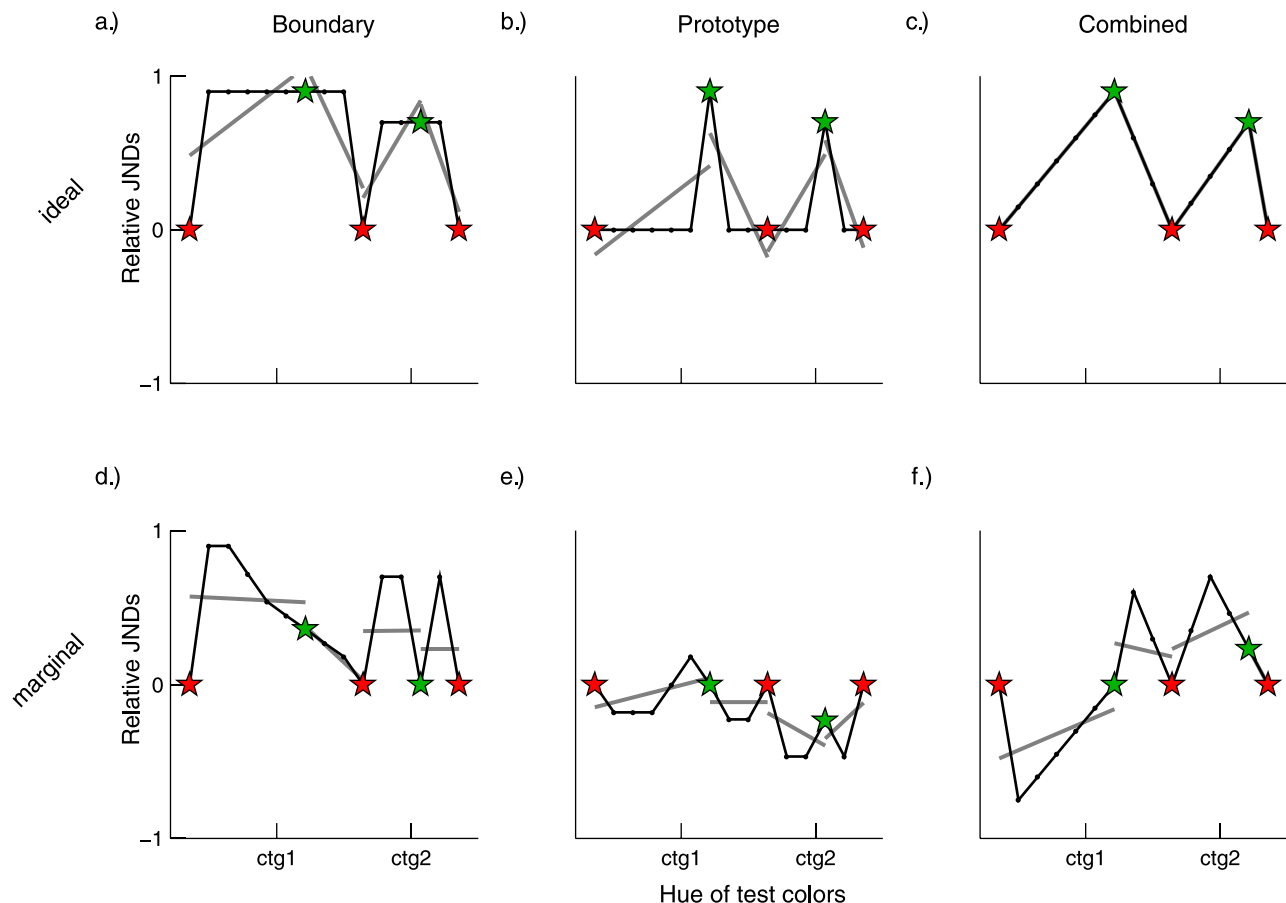


Figure 10. Models of categorical patterns. These graphics illustrate the different kinds of categorical patterns that may be expected in case of categorical sensitivity. The x-axis represents variation in hue between two example categories, ctg1 and ctg2. The y-axis corresponds to relative JNDs (for details see text and Figure 11). Black dots correspond to category JNDs, red stars to boundary JNDs, and green stars to prototype JNDs. The black line connecting the boundaries models the respective categorical pattern. Gray lines correspond to regression lines that show whether JNDs decrease (or increase) between prototypes and boundaries. The first row shows ideal models, which assume that category effects are the only determinants of JNDs. The second row illustrates marginal cases, in which the JND pattern is in line with one category effect, but contradicts the others. These cases can occur when JNDs are also modulated by other factors in addition to category effects. The columns correspond to a pure boundary effect, a pure prototype effect, and a combination of boundary and prototype effects, which results in a triangle-shaped pattern (panel c). *Our categorical perception tests were aimed to detect any of these category effects, even if JNDs are also modulated by other determinants.*

as thick black dots in Figure 9. The fact that these dots do not all have the same height reflects the impact of factors on the JND profile beyond the categories. As a result, the relationship between category and boundary JNDs is not reflected by the absolute size of the JNDs. Instead, this relationship is captured by the distances of the category JNDs from the line that connects the two boundary JNDs of each category (*boundary line*). These are the lines that connect the thick black dots in Figure 9. We will refer to the distances between JNDs and boundary lines as *relative JNDs* because they express JNDs relative to the boundary line. The different models in Figure 10 also use relative JNDs. For this reason the boundary JNDs in this figure (red stars) correspond to a value of zero along the y-axis.

However, additional noncategorical modulations of JNDs may also affect the JND pattern within each category. Furthermore, measurements will add noise to the JND pattern (cf. “Measurements” section). For these reasons, categorical patterns do not necessarily look as pure as illustrated in the first row of Figure 10. Instead, they may consist in local tendencies that are added to the global pattern and noise.

The combination with other sources of variability implies that not all kinds of category effects result in lower (relative) JNDs at the boundaries. Instead, JND patterns may agree with one kind of category effect, but contradict any other. This is illustrated by the patterns in the second row of Figure 10. Figure 10d shows examples of JND patterns, which are in line with boundary, but contradict prototype and triangle

effects. In particular, boundary JNDs of both categories (ctg1 and ctg2) in this panel tend to be lower than their category JNDs. At the same time, the prototype JND is not larger than the average category JND, and the regression line (in gray) has a slope of zero. The pattern in panel e, Figure 10 is in line with prototype effects, but contradicts triangle and boundary effects. Because prototype JNDs are locally protruding, they are higher than the average category JNDs. At the same time, boundary JNDs are not lower than category JNDs and the gray regression lines do not consistently increase towards the prototypes.

Finally, panel f, Figure 10 shows examples in which JNDs tend to decrease towards the boundaries even though boundary JNDs are not lower (ctg1) and prototype JNDs (ctg1 and ctg2) not higher than category JNDs. For example, the patterns in panel f may occur because the left boundary of ctg1 is slightly shifted away from the JND minimum, and the prototype of ctg2 is shifted away from the maximum. Such patterns may result when the boundaries and prototypes do not exactly coincide with JND minima or maxima, respectively. Given possible interactions with other determinants of JNDs and categories, a general tendency of JNDs to decrease towards the boundary may still indicate a category effect.

However, the JND pattern of ctg1 in panel f is ambiguous. It is not clear whether it is really due to a category effect of slightly shifted boundaries and prototypes. It could also reflect a pattern that is independent of the categories, and still yields increasing regression lines.

Categorical perception tests

Tests for categorical perception should be sensitive and specific to the different kinds of categorical patterns even if they are integrated into a global pattern of JNDs that is due to other factors. On the one hand, the tests should not miss a category effect because they are only sensitive to one kind of categorical pattern. On the other hand, the tests should not indicate a categorical pattern if it does not correspond to a category effect. In addition, it would be desirable that the tests distinguish between the different kinds of category effects.

For these reasons, we conceived three categorical perception tests. These tests compared category JNDs to boundary and prototype JNDs. We applied these tests to the relative JNDs to account for global modulations of JNDs that are not due to category effects. Figure 11 illustrates these relative JNDs for the aggregated data of Figure 9 (see Supplementary Figure S11 for other lightness levels). The three tests apply the models in the first row of Figure 10 to the relative JNDs.

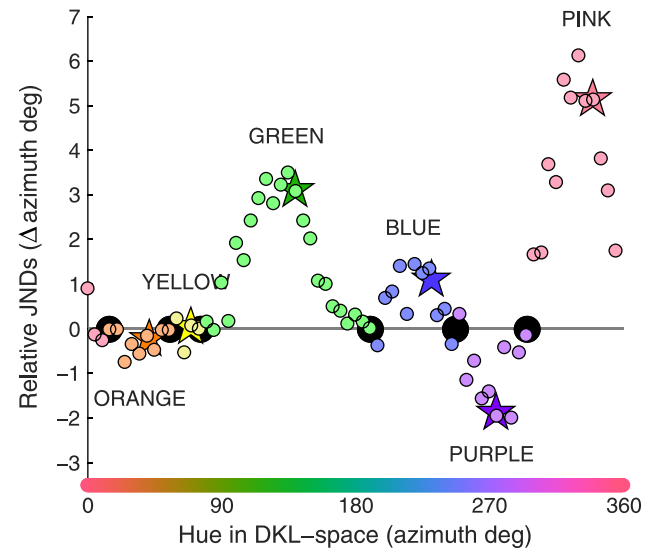


Figure 11. Relative JNDs. This graphic shows the relative JNDs for the aggregated data in Figure 9a. Relative JNDs are the distances of the category JNDs from the boundary lines that connect the black dots in Figure 9. The relative JNDs are shown along the y-axis. Format of the x-axis is as in previous figures. As in Figure 9, boundary JNDs are shown as black dots, which indicate the category boundaries in this figure. Colored discs represent single data points, and pentagrams the category prototypes. Their colors refer to the categories, which are also indicated by the color terms. This figure concentrates on results with the isoluminant background; see Supplementary Figure S11 for other lightness levels. *The relative JNDs enable statistical assessment of the categorical patterns of JNDs visible in Figure 9; compare for example, the hill-shaped pattern of green, blue, and pink in the corresponding graphics of the figures.*

First, in case of a boundary effect, relative JNDs should show a hill-shaped pattern between the boundaries (cf. Figure 10a and d). Hence, apart from statistical noise most category JNDs should lie above the boundary line (*boundary test*). In a first approach, we tested with a *t* test whether relative category JNDs were significantly above 0. The average across the relative category JNDs gives an impression of the size of the categorical effect.

However, the average tendency may be distorted by particular values of single JNDs. For example, a few extremely high JNDs can distort the average JND into a direction opposite to the majority of JNDs. For this reason, we also tested whether there were significantly more category JNDs above than below the boundary line. For this purpose, a second kind of boundary test evaluated through a binomial distribution whether the relative frequency of relative category JNDs above zero was significantly higher than 0.5.

Second, in case of a prototype effect the prototype JNDs should be higher than the category JNDs (cf. Figure 10b and e). We therefore tested through a *t* test

whether the (relative) category JNDs were lower than the (relative) prototype JND (*prototype test*).

Finally, a *triangle test* assessed statistical trends of JNDs to decrease towards the boundaries and to increase towards the prototypes (cf. Figure 10c and f). Such statistical trends of JNDs may be captured by a regression line for the relative JNDs between a boundary and the respective prototype. This was done for the values between each boundary and the adjacent prototypes, including not only category, but also boundary and prototype JNDs. In case of a category effect, the regression lines should form the two sides of a triangle with a downward tip around each boundary and of a triangle with an upward tip around each prototype.

For this reason, we calculated correlations between the relative JNDs and the distances of the test colors from the boundaries. This was done for the values between each boundary and the adjacent prototypes. This test resulted in a correlation coefficient for either azimuth side of the prototypes (e.g., left and right to the green prototype in Figure 9a). Positive correlation coefficients indicate that relative JNDs increase towards the prototype (relative to the boundary line) as predicted by a category effect.

These three tests complement each other. Each is particularly sensitive to one of the categorical patterns. The boundary tests can reveal the patterns in Figure 10d, but not those in Figure 10e, and vice versa for prototype tests. Moreover, boundary and prototype tests are insensitive to the distribution of the JNDs within the category. For this reason, the boundary tests are insensitive to the pattern of *ctg1*, and the prototype tests are insensitive to the pattern of *ctg2* in Figure 10f.

In contrast, the triangle tests are sensitive to the relative distribution of the category JNDs but not to the absolute sizes of the JNDs. For this reason, they indicate when category JNDs tend to increase towards the prototype even if they lie on the average below the boundary line (*ctg1* of Figure 10f), or above the prototype JND (*ctg2* of Figure 10f). This implies that they reveal such general tendencies even if the JNDs exactly at the boundary or at the prototype are not in line with a categorical pattern. At the same time, the triangle tests fail to show boundary or prototype effects without any clear tendency of JNDs to increase towards the prototype (Figure 10d, e).

Together, these tests allow us to detect any of the three kinds of categorical patterns, even if they are embedded in a global JND pattern and noise. In combination, they verify whether the JND patterns genuinely reflect category effects, and to specify which kind of category effect occurs. If JNDs are lowest at the boundary and increase towards the prototype, then all tests will be positive, indicating a particularly strong and clear category effect (as in Figure 9c). In contrast,

the absence of a categorical pattern in all of these tests indicates that there is no tendency towards any kind of category effect.

The statistics of all these tests were one-tailed. However, we also report significant two-tailed statistics when results contradict category effects. This is done in order to show that failures to reveal statistically significant category effects were not due to low statistical power.

Aggregated JNDs and consensus categories

For the analysis of the aggregated data, we applied the categorical perception tests to the data shown in Figure 9. Aggregated discrimination thresholds correspond to those shown in the polar plot of Figure 4. Altogether, there were 72 aggregated JNDs for isoluminant (Figure 9a) and dark colors (Figure 9b), and 32 JNDs for light colors (Figure 9c).

Aggregated categories should exclude idiosyncratic patterns of categorization, and represent the consensus across individuals. For this reason, we determined the boundaries of these *consensus categories* based on the mode over all single measurements of all $n = 10$ participants (cf. Figure 6 and Supplementary Figure S6). Prototypes are averaged across the respective $n = 9$ participants (cf. “Combination” section and lines in Figure 7 and Supplementary Figure S7). The resulting relative JNDs used in the categorical perception tests are shown in Figure 11 and Supplementary Figure S11.

Isoluminant colors

Above we reported three major dips and peaks for the isoluminant colors (“Discrimination” section). These three valleys were not local but extensive in that they gradually transitioned to hills in between the valleys (cf. Figure 9a). In line with a category effect, the local minimum at about 195° coincided almost exactly with the average green-blue boundary (191°), and the global minimum at 25° was close to the average pink-orange boundary (15°). Moreover, the peaks within the green ($\sim 120^\circ$ – 140°), and the blue category ($\sim 230^\circ$) were close to the green (140°) and the blue (231°) prototypes, respectively. The peak in the pink category (320°) was also close, but not coincident with the pink prototype (340°).

The category JNDs of pink, green, and blue formed a shape similar to a triangle and were above the boundary line (cf. Figure 11). This is captured by the results of the categorical perception tests: The boundary tests showed that category JNDs were higher than boundary JNDs for pink by on average 3.1° , $t(14) = 5.8$, $p < 0.001$, for green by on average 1.5° , $t(22) = 5.7$, $p < 0.001$, and for blue by on average 0.68° , $t(10) = 3.4$, $p =$

0.004. For all these categories most JNDs were above the boundary line, namely 87% for pink ($n = 15$, $p = 0.002$), 96% for green ($n = 23$, $p < 0.001$), and 82% for blue ($n = 11$, $p = 0.01$). Moreover, the prototype test confirmed that the prototype JNDs for pink, green, and blue were still further away from the boundary line than the category JNDs, namely by 2.0° , $t(14) = 3.8$, $p = 0.001$; 1.6° , $t(22) = 6.2$, $p < 0.001$; and 0.4° , $t(10) = 2.1$, $p = 0.03$, respectively. Finally, triangle tests were significant on both sides of the prototypes for pink ($n = 8$, $r = 0.89$, $p = 0.001$ and $n = 7$, $r = 0.99$, $p < 0.001$), and green ($n = 12$, $r = 0.94$ and $n = 11$, $r = 0.92$; both $ps < 0.001$). For blue, it was significant on the lower-azimuth side ($n = 8$, $r = 0.71$, $p = 0.02$), but did not reach significance on the other side because the prototype was close to the boundary (cf. Figure 11) and there were too few measurements in between ($n = 3$, $r = 0.77$, $p = 0.22$).

However, for orange and purple, the results contradicted the categorical pattern. Their category JNDs were on average below the boundary line by 0.25° , $t(8) = -2.7$, $p = 0.03$, and 0.95° , $t(9) = -3.8$, $p = 0.004$; most category JNDs were below the boundary line, namely 100% ($n = 9$, $p = 0.002$) and 90% ($n = 10$, $p = 0.01$); and the prototype JND of purple was still -0.94° further below the boundary line than the category JNDs, $t(9) = -3.8$, $p = 0.004$. The orange prototype lay exactly on the boundary, mean = 0.01° , $t(8) = 0.12$, $p = 0.91$. For the aggregated data, tests were not sensible for yellow due to the small number of measurements within the narrow yellow category ($n = 4$; cf. Figure 11).

Taking all $n = 72$ JND measurements together, they were on average above the respective boundary lines by 1.1° , $t(71) = 4.9$, $p < 0.001$, and more frequently above the boundary lines (67%, $n = 72$, $p = 0.001$), but not significantly lower than the average prototype JND at 0.15° , $t(71) = 0.68$, $p = 0.25$.

Dark colors

In the condition with the white background (cf. Figure 9b), the minimum at 15° was close to, but not coincident, with the red-brown boundary (29°). Moreover, the local minimum at 190° was also close to the green-blue boundary of this lightness level (200°). However, it is questionable whether it is a category-specific dip since this minimum is part of an overall valley that extends from green to purple, hence covering the whole blue category.

The categorical perception tests all confirmed a categorical pattern for green (cf. Supplementary Figure S11a). Its category JNDs were on average 1.8° , $t(26) = 6.2$, $p < 0.001$, and with 93% most frequently above the boundary line ($n = 27$, $p < 0.001$). The prototype JND was 1.1° further away from the boundary line than the category JNDs, $t(26) = 3.8$, $p < 0.001$, and both correlations in the triangle test were positive and

significant ($n = 15$, $r = 0.68$, $p = 0.003$ and $n = 12$, $r = 0.86$, $p < 0.001$). The category JNDs of purple were on average 0.8° , $t(17) = 2.8$, $p = 0.006$, and 67% ($n = 18$, $p = 0.04$) above the boundary line in the boundary tests, but the triangle test failed to show a triangular pattern for purple ($n = 6$, $r = 0.22$, $p = 0.34$ and $n = 12$, $r = -0.35$, $p = 0.26$).

In contrast, the category JNDs of red lay by 1.1° , $t(5) = -3.8$, $p = 0.01$, and with 100% ($n = 6$, $p = 0.02$) below the boundary line in the boundary tests, and the prototype JND of red was marginally significantly lower than the category JNDs (0.76° , $t(5) = 2.6$, $p = 0.05$). Moreover, the distances from the boundary line of the blue prototype JNDs were lower than those of the category JNDs (mean = 0.49° , $t(8) = 2.4$, $p < 0.05$). All other tests did not reach significance.

The tests across all $n = 72$ JNDs yielded similar results with the dark as with the isoluminant colors: All JNDs together were on average 0.8° , $t(71) = 4.5$, $p < 0.001$, and most frequently above the boundary line (65%, $n = 72$, $p = 0.002$). However, the average category JND was also significantly higher than the average prototype JND (-0.7° , $t(71) = 4.1$, $p < 0.001$), which contradicts a prototype effect.

Light colors

Finally, there were still fewer categorical patterns for light colors (cf. Figure 9c). There was no dip at the green-blue boundary (192°) at all. Instead, one of the troughs (30° – 60°) occurred around the orange-yellow boundary (53°). However, the other two dips (150° and 270°) were clearly located within the green and the purple category, and not close to any category boundary. Apart from these overall trends, there were no local dips at the boundaries or peaks at the prototypes for the aggregated data.

The categorical perception tests confirmed that most JND patterns were incompatible with category effects (cf. Supplementary Figure S11b). The only exceptions were the pink prototype that was further away from the boundary line than the pink category JNDs (mean = 0.94° , $t(7) = 2.35$, $p = 0.02$), and the triangle tests on the lower-azimuth side of blue ($n = 4$, $r = 0.96$, $p = 0.02$; there were not enough measurements on the other side).

In contrast, the categorical perception tests revealed JND patterns that contradicted any category effect. In particular, green yielded results opposite to the categorical pattern that were significant in the prototype (-4.7° , $t(10) = 4.12$, $p = 0.002$) and triangle tests ($n = 6$, $r = -0.91$, $p = 0.01$ and $n = 5$, $r = -0.96$, $p < 0.01$), and marginally significant in the two boundary tests (-2.5° , $t(10) = -2.2$, $p = 0.05$, and 73%, $n = 11$, $p = 0.08$, respectively). Tests for other single categories were not sensible because they contained too few JND measurements for light colors (orange: 4, yellow: 3, blue: 5,

and purple: 5; cf. Figure 9c and Supplementary Figure S11b).

However, the overall tests showed that the JNDs for light colors tended into the direction opposite to the category effect. Category JNDs were on average -1.0° below the boundary line, $t(35) = -2.2$, $p = 0.03$. The other boundary test was not significant (44%, $n = 36$, $p = 0.11$). However, the prototype JND was still lower than the category JNDs (-0.90° , $t(35) = 2.0$, $p < 0.05$), which contradicts a prototype effect.

Individual JNDs and categories

Due to the individual differences in color naming, the JNDs per category will involve different hues, different numbers of data points, and different distances between boundaries and prototypes across individuals. For this reason, group statistics can only be applied to the individual data after establishing categorical patterns for each individual separately.

In a first step we applied the categorical perception tests to each individual dataset shown in Figure 8 and Supplementary Figure S8. The individual relative JNDs used in these tests are provided in Supplementary Figures S9 and S10. Note that for dark and light colors, discrimination thresholds, but not categories, were interpolated for six and eight observers, respectively, and the number of individual datasets per category varies depending on the respective category sets (cf. “Measurements” section).

Figure 12 illustrates the results of the individual boundary tests using the example of isoluminant colors (see Supplementary Figure S12 for other lightness levels). Bars correspond to the average relative JNDs. Bars above zero are in line with boundary effects. Similar graphics for the other categorical perception tests are provided in Supplementary Figure S13.

In the second step, we applied paired t tests across individuals. Hence, for color names that were used by all participants the sample size was $n = 10$; for color names that were only used by some of the participants, it was $n < 10$. To test for boundary effects, the t tests assessed whether the relative JNDs (bars in Figure 12 and Supplementary Figure S12) were on average significantly larger than zero. Moreover, we tested whether the frequency of JNDs above the boundary line (bars in Supplementary Figure S13a through c) was higher than predicted by chance ($p = 0.5$). For prototype effects we tested whether the differences between prototype and category JNDs (bars in Supplementary Figure S13d through f) were above zero. Finally, we assessed whether correlation coefficients on both sides of the prototypes (both kinds of bars in Supplementary Figure S13g through j) were

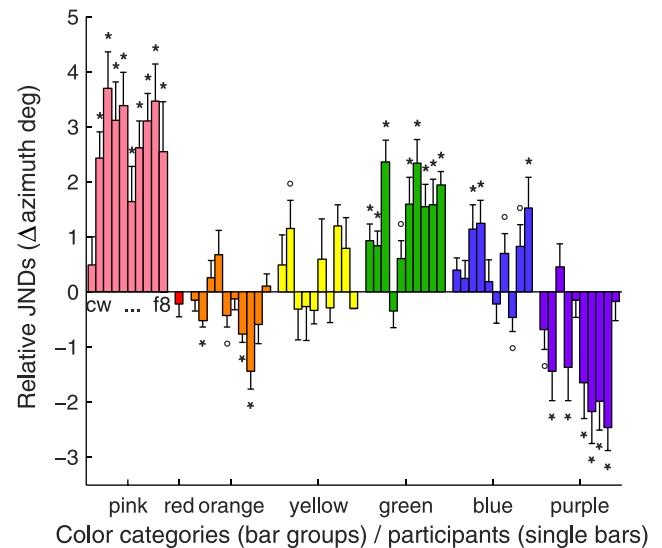


Figure 12. Individual boundary tests for averages. The graphic illustrates the single boundary tests for average category JNDs using the individual data with isoluminant colors shown in Figure 8 (see Supplementary Figure S12 for other lightness levels). Bars represent the average difference between category and boundary JNDs. Each bar refers to a test for one individual and one category. Categories are listed along the x-axis and indicated by the bar colors. The bars for each participant follow the order of participants as in Figure 8 (CW to F8); the single bar for the red category refers to F8. Error bars show SEM, symbols indicate p values (* for $p < 0.05$, ° for $p < 0.1$). Degrees of freedom correspond to $n - 1$ of category JNDs (for n see Figure S13a). Note that the average category JNDs of pink, green, and blue were for most participants higher, those of purple lower than the boundary line.

positive. For this purpose, we applied t tests to the Fisher-transformed correlation coefficients.

Figure 13 gives an overview of these tests across individuals. In general, this figure is relatively easy to interpret: The higher the values along the y-axis, the more the respective JND pattern is categorical.

The boundary test is illustrated by the bars in this figure. The height of the bars represents the average difference between category and boundary JNDs, and corresponds to the average across the bars in Figure 12 and Supplementary Figure S12, respectively. The higher the bars are above zero, the higher the relative category JNDs are.

The boundary test for frequencies is illustrated by the relation of saturated and unsaturated areas within each bar (see also Supplementary Figure S13a through c). The higher the proportion of the saturated area the more JNDs lay consistently above or below the boundary line. For significant differences from 0.5, the relative frequency, is provided at the base of the bars. In sum, the higher and the more saturated the bars are, the more they are in line with a boundary effect.

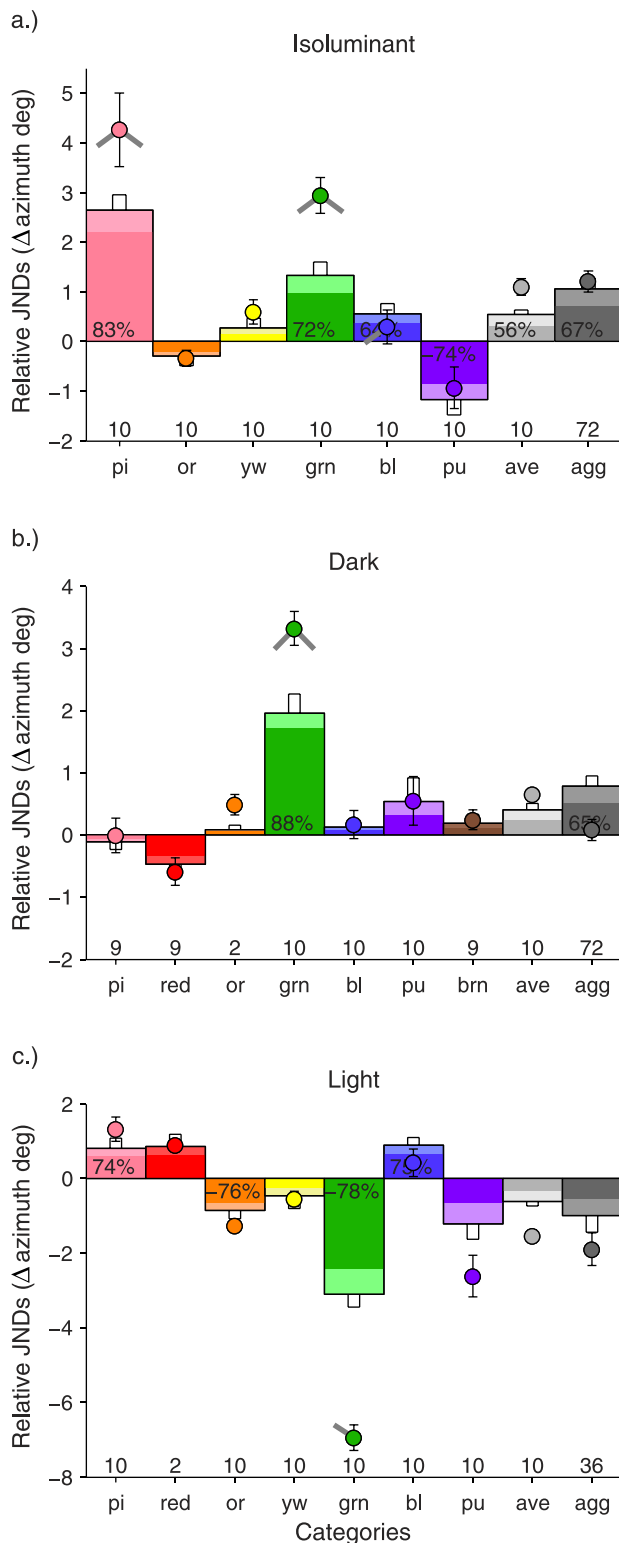


Figure 13. Tests across individuals. These graphics illustrate the categorical perception tests across individuals for isoluminant (panel a), dark (panel b), and light colors (panel c). Each colored bar corresponds to a category, the light gray bar to the average across categories, and the dark gray bar to the overall tendency of the aggregated data (cf. Figure 9). The bars illustrate the boundary tests. Their height reflects the average difference

The results of the prototype test are illustrated by the colored disks. They correspond to the average of the bars in Supplementary Figure S13d through f. The higher a colored disk is above the corresponding bar, the more the prototype JND was above the average category JND, as predicted by the prototype effect.

Finally, the results of the triangle tests are illustrated by the tilted gray lines around the colored disks. These lines are only shown for significant correlation coefficients. If JNDs increase on both sides of the prototype, the lines form the tip of a triangle pointing upwards (cf. pink and green in Figure 13a). This confirms a triangle-shaped categorical pattern around the prototype and points towards a combination of boundary and prototype effects.

If all criteria of categorical perception hold, the bars should be significantly larger than zero, the saturated areas should be significantly larger than the unsaturated areas, the disks should be significantly larger than the bars, and the gray lines should form a triangle that points towards the disks. For example, this is the case for isoluminant pink and green and for dark green.

In order to appreciate the overall tendency, we calculated for each participant the average across categories. For example, for the overall boundary test we averaged the bars in Figure 12 across categories. We applied t tests across participants to these average values. The results are illustrated by the light gray bars in Figure 13. For comparison, the dark gray bars in Figure 13 illustrate the tendencies across all categories found with the aggregated data, as reported above (“Aggregated JNDs and consensus categories” section).

Isoluminant colors

At isoluminance, only the average category JND for pink, $t(9) = 8.6$, $p < 0.001$; green, $t(9) = 5.0$, $p < 0.001$;

between category and boundary JNDs. The small white bars on top show SEM . The repartition of each bar into a saturated and unsaturated area reflects the amount of JNDs above and below the boundary line. When this amount is significantly different from chance, the percentage is shown at the base of the bar. The colored disks indicate the average difference between category and prototype JNDs (prototype test). Error bars depict SEM . Finally, the tilted gray lines are shown when the correlation coefficients of the triangle tests are different from zero in a two-tailed t test with $p < 0.05$. The number of individual datasets per category is shown above the x-axis. For the aggregated JNDs (dark gray bar) this number corresponds to the number of test colors, for which JNDs were measured. Overall the individual data only mirrors the patterns found for the aggregated data, and does not show clear categorical patterns apart from those for isoluminant pink, green, and blue, and dark green (cf. Figure 9).

and blue, $t(9) = 2.7$, $p = 0.01$, lay above the respective boundary line (cf. bars in Figure 13a). Moreover, for these categories the frequency of category JNDs above the boundary line was above 50%, namely for pink with 83%, $t(9) = 8.1$, $p < 0.001$; for green with 72%, $t(9) = 3.4$, $p = 0.004$; and for blue with 64%, $t(9) = 2.4$, $p = 0.02$, respectively. For pink and for green, prototypes also yielded higher JNDs than the average category JNDs, $t(9) = 2.2$, $p = 0.03$, and $t(9) = 4.4$, $p < 0.001$, respectively. Finally the triangle tests indicated a triangle-shaped distribution around the prototypes of these categories. JNDs decreased on both sides of the prototypes for pink, $t(9) = 4.4$ and 6.6 , both $ps < 0.002$, and green, $t(9) = 7.1$ and 8.0 , both $ps < 0.001$.

In contrast, the boundary tests for purple confirmed that category JNDs lay on average, $(-1.2^\circ, t(9) = -3.8, p < 0.005)$, and more frequently below the boundary line (74%, $t(9) = -3.8, p < 0.005$). Category JNDs for orange also contradicted the category effect (-0.3° and 67% below the boundary line); but the respective boundary tests did not reach significance with two-tailed statistics, $t(9) = -1.6$, $p = 0.15$, and $t(9) = -2.2$, $p = 0.06$.

The overall boundary tests showed that category JNDs in general were significantly higher than boundary JNDs (light gray bar in Figure 13a). Relative category JNDs were on average 0.54° and in 56% of the measurements above zero, $t(9) = 5.8$, $p < 0.001$, and $t(9) = 2.8$, $p = 0.01$. Finally, the average prototype JND was 0.56° higher than the average category JND, which was also significant, $t(9) = 3.4$, $p < 0.004$.

Dark colors

For dark colors (cf. Figure 13b), only green showed the expected pattern with the average category JND, $(2.0^\circ, t(9) = 6.3, p < 0.001)$, and more than 50% of category JNDs above the boundary line (88%, $t(9) = 13.1, p < 0.001$), the prototype JND above the category JNDs, $(1.4^\circ, t(9) = 5.0, p < 0.001)$, and correlations that indicate a triangle-shaped pattern, $t(9) = 7.2$ and $t(9) = 8.7$, both $ps < 0.001$.

As with isoluminant colors, the overall tests showed general tendencies towards the categorical pattern (cf. light gray bar in Figure 13b). When aggregated across categories, category JNDs were on average higher than boundary JNDs by 0.4° , $t(9) = 3.8$, $p = 0.002$; over 50% of them lay above the boundary line, namely 57%, $t(9) = 2.0$, $p = 0.04$, and prototype JNDs were still higher by 0.26° than category JNDs, $t(9) = 2.6$, $p < 0.02$.

Light colors

For light colors (cf. Figure 13c), the pink and blue relative category JNDs were on average, $(0.8^\circ, t(9) = 3.1, p = 0.007)$, and $(0.9^\circ, t(9) = 4.0, p = 0.002)$, and more often than chance above zero, namely 74%, $t(9) = 3.1$, $p = 0.007$, and 75%, $t(9) = 3.9$, $p = 0.002$. In contrast,

orange, green, and purple showed the reverse pattern. Their category JNDs were lower than average boundary JNDs: -0.9° , $t(9) = -3.7$, $p = 0.005$; -3.10° , $t(9) = -9.0$, $p < 0.001$; and -1.2° , $t(9) = -3.0$, $p = 0.02$, respectively, and those of orange and green were mostly below the boundary line: 76%, $t(9) = -4.4$, $p = 0.002$; and 78%, $t(9) = -18.4$, $p < 0.001$.

Unlike the overall tests for isoluminant and dark colors, these tests contradicted any categorical pattern for light colors (cf. light bar in Figure 13c). Category JNDs were below the average boundary JND by -0.6° , $t(9) = -5.1$, $p = 0.001$, and the relative frequency of the category JNDs above the boundary line was not significantly different from chance level at 48%, $t(9) = -0.7$, $p = 0.53$. Finally, overall the prototype JNDs were still lower than the category JNDs at -0.9° , $t(9) = -5.5$, $p < 0.001$.

Discussion

According to the categorical perception hypothesis, there should be local minima of JNDs around the category border, local maxima around the typical colors, and/or a tendency of JNDs to increase from the boundaries towards the prototypes. These patterns should be specific to the location of the categories that correspond to the basic color terms.

In the “Results” section we applied the different categorical perception tests to all categories, at all lightness levels, and to individual and aggregated data in order to make sure that we did not miss any traces of potential category effects. We will only conclude about the absence of a categorical pattern when there was not one such effect in any of the categorical perception tests. In contrast, when concluding about the existence of a systematic pattern, the interpretation of significant results in multiple tests may yield spurious effects due to unsystematic statistical variation. If we consider the single tests for each category as subtests of one overall test, a correction for multiple testing is needed. The Bonferroni correction requires the multiplication of p values by the number of categories ($n = 8$ for isoluminant and light, $n = 6$ for dark colors). Moreover, we will only interpret results that were significant in all categorical perception tests. Results that were consistent across tests, but did not reach significance in some of the tests, are considered as descriptive results, only.

Category effects in DKL space

At all lightness levels, JNDs followed an overall pattern across hues (cf. “Discrimination” section). However, the profile of the aggregated JNDs seems not to be category-specific (cf. Figure 9). While some dips

and peaks of JNDs were congruent with the categorical perception hypothesis, others were not. The results for the individual data sets mirrored the overall pattern of the aggregated data; they did not reveal additional categorical patterns at the individual level (cf. Figure 13 to Figure 9).

On the one hand, we observed categorical patterns for pink, green, and blue with isoluminant, and for green with dark colors (cf. Figure 13a, b). Even with the conservative Bonferroni correction, all tests were still significant for isoluminant pink, green, and dark green since all the respective uncorrected p values were below 0.00625 (i.e., $1/8$ of 0.05). Overall boundary tests across all categories were in line with a general tendency towards a categorical pattern for isoluminant and dark colors (gray bars in Figure 13a, b).

On the other hand, isoluminant orange and yellow did not show any tendency towards a category effect, and purple contradicted any category effect (cf. Figures 9a and 13a). Moreover, reducing the lightness of the stimulus colors strongly changed categorization, but barely affected color sensitivity (cf. “Measurements” section). This implies that categorization changed independently from sensitivity. As a result, there were still fewer categorical patterns with dark than with isoluminant colors (cf. Figures 9b and 13b). These results indicate that category changes across lightness are not coupled to corresponding changes in color sensitivity. For light colors, discrimination thresholds changed strongly, but categories did not change accordingly (cf. “Measurements” section). Consequently, JNDs for light colors completely contradicted any category effect (cf. Figures 9c and 13c). These results show that categorization is not affected by changes in sensitivity due to adaptation.

In sum, the troughs and peaks of the JND profile were not localized specifically at the category borders and prototypes. Instead, the JND profile consisted of global changes in JNDs across hues (cf. Figure 9). These global patterns of JNDs coarsely coincided with categories in some regions of color space (green, pink, and maybe blue); in other regions they did not (yellow, orange, and in particular, purple). This concordance between the global pattern of JNDs and some of the categories was strong enough to yield an overall categorical pattern with the isoluminant and dark colors. In these cases opposite effects in the other categories were too weak to counteract the strong categorical patterns of those categories (cf. size of positive and negative bars in Figure 13a, b).

We conclude that the coincidence of some boundaries with JND minima is not a general property of the categories that correspond to the basic color terms. For this reason, the few categorical patterns we observed cannot reflect genuine effects of the linguistic categories on color sensitivity.

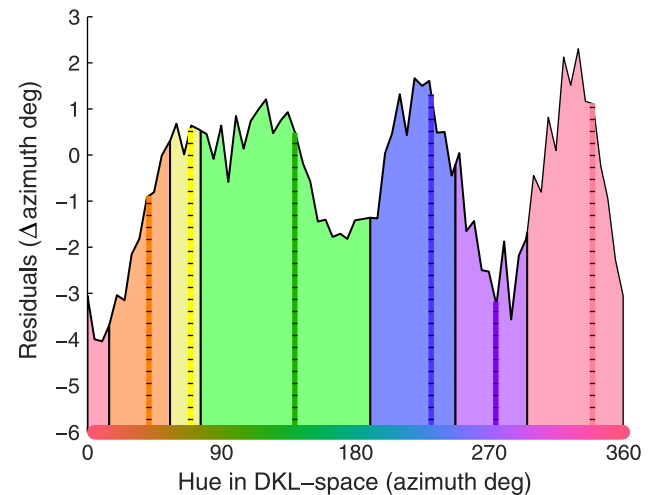


Figure 14. Residuals of ellipse fit. The curve reproduces the residuals of Figure 5b to compare them with the categories. Format as in Figure 9a. This figure shows the residuals for isoluminant colors, corresponding figures for other lightness levels are provided in Supplementary Figure S14. Categorical patterns occur for the same categories as with the original JNDs, namely for green, blue, and pink, but not for orange, yellow, and purple.

Categorical sensitivity and perceptual mechanisms

At the same time, the strong concordance of some categories with the global JND pattern might imply that the sensitivity to color differences has some impact on color categorization. The global pattern of JNDs is most probably shaped by basic perceptual mechanisms, such as the second-stage and cortical mechanisms of color vision. Above we observed that the most general pattern of JNDs followed an elliptical shape, at least for isoluminant and dark colors. For isoluminant colors, we also found that the residual variation of the JNDs is most probably modulated by the second-stage mechanisms (cf. “Discrimination” section). Consequently, JNDs for isoluminant colors show systematic deviations from the smooth transition represented by the ellipse. In this sense, the JND pattern for isoluminant colors is categorical. The question arises whether this categorical pattern is related to the linguistic color categories.

Figure 14 compares the residuals of the fitted ellipse for the isoluminant colors with the corresponding color categories. The minima of these residuals align more clearly with the DKL-axes than the original JNDs (cf. Figure 5). Nevertheless, the three major dips are still close to the pink-orange and green-blue boundary and to the purple prototype. The categorical pattern for green is less pronounced for the residuals than for the original JNDs. This is due to the fact that the center of the green category is located around the vertex of

maximal curvature of the ellipse. As a result the ellipse accounts for the higher JNDs in green. At the same time, blue shows a more pronounced categorical pattern with the residuals. Apart from green and blue, the categorical pattern for pink reappears, and orange, yellow, and purple still disagree with any categorical pattern. Overall, similar categorical patterns appear when accounting for the most general JND pattern represented by the ellipse. Since the pattern of these residuals most probably reflects the impact of the second-stage mechanisms this finding establishes a link between color categorization and second-stage mechanisms.

For dark and light colors, there was little evidence for the impact of the second-stage mechanisms. The JNDs for dark colors basically follow the global ellipse (cf. Figure S5a, b). When accounting for this elliptical shape the categorical pattern for green disappears, but new potentially categorical patterns appear for brown and blue (cf. Supplementary Figure S14a). This observation, together with the findings for isoluminant colors, indicates that the strong categorical pattern for green somehow depends on the elliptical shape. For light colors, the inspection of the residuals yields the same findings as with the original JNDs: The residuals for light colors completely contradict any categorical pattern (cf. Figure 14b).

Overall, the concordance of some categories with the global JND pattern points towards a loose relationship between color categorization and the perceptual mechanisms of color vision. In this regard, sensitivity to color differences may be considered as categorical in the broad sense. However, the global JND pattern is not specific to the categories that correspond to the basic color terms. For this reason, our basic ability to discriminate colors cannot fully explain why we use this particular set of categories to communicate about colors. In regard to these linguistic categories, sensitivity for color differences is not categorical at the level of color processing represented by the DKL space.

Generalization to other color spaces

The observations above are also true for other color spaces. CIELAB and CIELUV have been used in previous studies to represent color perception at higher levels of processing. We recalculated the JNDs of the present study as Euclidean distances in CIELUV and CIELAB. For this purpose, we converted the original test and comparison colors at the reversal points into the respective color space. Then we calculated Euclidean distances between these colors at each reversal point, and averaged the resulting distances across reversal points to obtain JNDs.

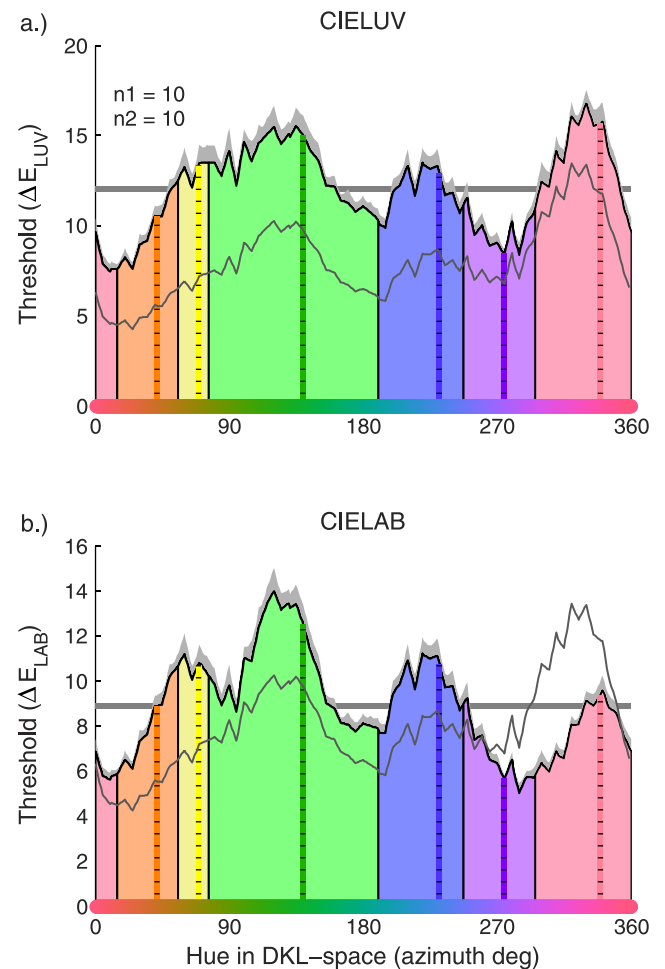


Figure 15. JNDs in CIELUV and CIELAB. Aggregated JNDs in CIELUV (panel a) and CIELAB (panel b) are shown for isoluminant colors. Format is the same as in Figure 9. JNDs are calculated as Euclidean distances. The y-axis represents these Euclidean distances. Note that for comparison the x-axis is the same as in Figure 9 and represents hue as azimuth in DKL space. The dark gray line reproduces the JND curve in DKL space from Figure 9a. As in DKL space, green, blue, and pink, but not orange and purple, are in line with a categorical pattern in both spaces.

Figure 15 shows the aggregated JNDs for the isoluminant colors in CIELUV (panel a) and CIELAB (panel b). For comparison, hues vary along the x-axis as a function of azimuth in DKL space. The gray curves reproduce the JNDs in DKL space from Figure 9a. Not surprisingly, the global pattern of JNDs differs across the color spaces. However, the local troughs and peaks are mostly the same as in DKL space. In CIELUV, there are the same three peaks and dips we observed in DKL space. In CIELAB, there is an additional fourth peak around the orange-yellow boundary. However, this peak contradicts a category effect because it coincides with a boundary instead of a prototype.

Supplementary Figure S15 illustrates the corresponding categorical perception tests across individuals. Like in DKL space, there was an overall tendency of category JNDs to lie above the boundary lines (cf. gray bars in Supplementary Figure S15). The respective overall boundary tests were highly significant in CIELUV (all $ps < 0.02$) and CIELAB (all $ps < 0.01$). In addition, prototype JNDs tended to be higher than category JNDs (cf. gray disks in Supplementary Figure S15).

Moreover, the JNDs for green, blue, and pink were triangle-shaped in both CIELUV and CIELAB. So, again these categories were completely in line with the patterns of a combined category and prototype effect. Orange and purple were not in line with any categorical pattern, as it was the case in DKL space. For yellow, the results were ambiguous. Yellow category JNDs tended to lie above the boundary line (cf. yellow bars in Supplementary Figure S15). The boundary tests on average category JNDs across individuals were significant in both, CIELUV: 0.67° , $t(9) = 1.8$, $p = 0.049$, and CIELAB: 0.7° , $t(9) = 2.7$, $p = 0.01$. However, the other categorical perception tests for the yellow category were not significant in any of the two spaces.

In sum, these observations underpin our conclusions above. The distribution of JNDs shows categorical patterns for some (green, blue, pink) but not for other categories (orange, purple). Hence, the JND patterns are not specific to the categories in CIELUV and CIELAB, either. We conclude more generally that the categories that correspond to the basic color terms are not inherent to color sensitivity.

Comparison to previous findings

Since we did not find local minima of discrimination thresholds at the category boundaries, our results seem to agree with those of Roberson and colleagues (2009). Like us, Roberson and colleagues did not find evidence for category effects on discrimination thresholds. However, there were also differences between our results. Roberson and colleagues measured discrimination thresholds for colors that cross the green-blue boundary and found that the thresholds were rather constant. In contrast, we found a pronounced trough in this region of the color space.

The fact that they represented colors in CIELUV space might be one reason why they did not find variations of JNDs across green-blue hues since CIELUV is more homogenous in terms of JNDs than DKL color space. However, our JNDs varied considerably in CIELUV (cf. Figure 15a). In particular, the trough at the green-blue boundary reappears not only in CIELUV, but also in CIELAB (cf. Figure 15b).

Moreover, in a previous study we measured discrimination thresholds for several rendered versions of Munsell Chips in CIELUV space (Witzel & Gegenfurtner, 2011). Although, the profile of JNDs also depended on the background to which participants adapted, we always found clear differences of JNDs in CIELUV space among colors in the green-blue and in the blue-purple region (cf. supplementary figures S5 and S6 in Witzel & Gegenfurtner, 2011). Taken together, these results show that there are clear differences in JNDs in CIELUV space, and that the absence of differences in the study of Roberson and colleagues is not just due to the representation of color differences in CIELUV.

Another reason for the differences between their and our results might be that they used a different method to measure discrimination thresholds. Their observers had to indicate the location of a border between two adjacent color areas. Sensitivity for color edges is much higher than for the color discrimination we measured in our study. As a result, their discrimination thresholds are so tiny that the corresponding color differences cannot even be rendered on a usual monitor with 8-bit color resolution (see figure 1c in Roberson et al., 2009). The strong difference in the size of JNDs shows that the discrimination thresholds found by Roberson and colleagues are qualitatively different from ours. This qualitative difference is most probably the reason for the difference between their and our profile of discrimination thresholds.

Moreover, our study provides an explanation of Regier, Kay, and Khetarpal's (2007) results beyond the insufficient control of perceptual differences. The statistical tendency towards a higher sensitivity across than within categories implies that, on average, there are higher color differences across than within categories, as claimed by Regier and Kay. This general tendency may be explained by the coarse concordance of some category boundaries with the JND minima, in particular those in line with the second-stage mechanisms (pink-orange and green-blue; cf. Figure 9). Under the assumption that sensitivity across hues is roughly stable across cultures, the loose relationship between categorization and sensitivity may also constitute the origin of the statistical patterns in categorization across cultures (Kay & Regier, 2003).

Furthermore, our results contradict the idea that categorization shapes color discrimination through perceptual learning (Özgen, 2004; Özgen & Davies, 2002). According to this idea perceptual learning occurs due to categorization when applying color names in everyday life. Perceptual learning would boost discrimination specifically at the location of each category boundary. Our findings show that this is not the case. Özgen and Davies (2002) used a delayed same-different task. Such a task does not measure pure discrimination

since it strongly involves attention and memory. Effects of perceptual learning as well as other kinds of category effects observed in previous studies may exist at a more cognitive level of color processing—for example due to effects on memory or attention. However, they do not exist for the basic ability to perceptually discriminate colors.

Finally, we did not observe that sensitivity increased towards focal yellow and blue, or any other hue around the centers of these categories. Hence, our results do not support the observation of Danilova and Mollon (2010, 2012). The lack of this effect might be due to differences in procedure, in particular the longer presentation time of the stimulus display in our experiments. However, further research is necessary to establish a link between their and our results.

The control of perceptual differences

Previous studies on categorical perception have controlled perceptual distances in terms of Munsell steps or Euclidean distances in CIELUV or CIELAB space (cf. Introduction). These measures, however, do not capture the fine-grained perceptual distances as necessary for the study of category effects.

Our results show that at certain lightness levels some of the categories roughly coincide with the strong variations of JNDs across the hue circle. As a consequence, it is possible that those studies confound global variations in sensitivity with category effects. In this case, the patterns assumed to be category effects might rather be due to the particularities of the stimulus sampling than to genuine category effects (see also Witzel & Gegenfurtner, 2011). This idea is further supported by studies that have shown that large differences of reaction times in visual search did not reflect color categories, but could rather be explained by the color opponent channels (Lindsey et al., 2010; A. M. Brown et al., 2011).

In this respect, the variation of JNDs around the green-blue boundary is particularly important. We observed that this category border coincides with the center of a pronounced trough of JNDs that occurred when participants adapted to an isoluminant gray background. At the same time, the green-blue boundary is close to the axis representing the (L – M) mechanism (cf. “Discrimination” section and Malkoc et al., 2005). Hence, the coincidence of the second-stage mechanism with the green-blue boundary might be the explanation for the wide trough of JNDs in the region around the green-blue boundary for isoluminant colors (e.g., Krauskopf & Gegenfurtner, 1992).

Many studies on categorical perception have used the green-blue boundary as the prime example to investigate categorical color perception (e.g., Bornstein

& Korda, 1984; Kay & Kempton, 1984; Witthoft et al., 2003; Gilbert et al., 2006; Drivonikou et al., 2007; Franklin, Drivonikou, Bevis, et al., 2008; Franklin, Drivonikou, Clifford, et al., 2008; Siok et al., 2009; Fonteneau & Davidoff, 2007; Holmes et al., 2009; Özgen & Davies, 2002; Roberson et al., 2009). These studies mostly presented stimuli as distinctly colored areas (disks or squares) on a gray background. This kind of stimulus display is more similar to ours than to the one of Roberson et al. (2009). Furthermore, the gray background corresponds to our condition with the isoluminant background.

As a result, this classical set of stimuli should produce the boost of sensitivity we found at the green-blue boundary with the isoluminant colors. And indeed, we observed such a pattern of sensitivity for these stimuli in a previous study (Witzel & Gegenfurtner, 2011). Since the (L – M) mechanism coincides with this boundary, it provides an alternative explanation for the increase of perceptual difference at this boundary. As a consequence, the effects observed in previous studies may actually be due to the second-stage mechanisms rather than to the impact of color categories (see also A. M. Brown et al., 2011). The results of these studies might well not replicate if other colors were studied, and differences in sensitivity were controlled.

Conclusion

We investigated the relationship between color sensitivity and color categorization. In particular, we tested whether there are category effects on discrimination thresholds. In fact, for isoluminant and dark colors we found a general tendency of sensitivity to be higher across than within categories. In particular, when averaged across all categories, JNDs tended to be lower at category borders and higher within categories. This finding validated ambiguous results of previous studies (Regier et al., 2007).

However, detailed analyses showed that the pattern of JNDs was not specific to the categories that correspond to the basic color terms. The overall tendency towards a categorical pattern was mainly due to the coarse concordance of some category borders and prototypes with the global pattern of JNDs (cf. Figure 9). When colors were isoluminant to the background, this was the case for the pink, the green, and to a lesser extent for the blue category (cf. Figure 13a). Lower lightness yielded a different set of categories, but no corresponding change in JNDs. As a result, green remained the only category that was in line with a categorical pattern at lower lightness (cf. Figure 13b). None of the other categories at both

lightness levels showed any category effect. In addition, the concordance between JNDs and the aforementioned categories was only coarse and unspecific (cf. Figure 9). Finally, the change of color sensitivity due to changes in adaptation did not imply a corresponding change of color categorization. Hence, when colors were shown on a black background JNDs completely contradicted any category effect (cf. Figure 13c). These results undermine the idea of categorical perception because the JND patterns are not specific to the categories.

The global pattern of JNDs coarsely followed an ellipse. The categorical pattern of isoluminant and dark green seems to be partly due to its location on the respective ellipses. For isoluminant colors, deviations from the ellipse coincided with the second-stage cone-opponent mechanisms. The orange-pink and green-blue boundaries roughly coincided with the reduction of JNDs around the second-stage mechanisms (cf. Figure 14). Hence, the second-stage mechanisms provide a possible explanation for the increase in sensitivity at these boundaries.

Together, our results point towards a loose relationship between color sensitivity and categorization. This relationship implies that the sensitivity to color differences may affect color categorization. It also points towards a potential link between color categories and the perceptual mechanisms that shape the global pattern of sensitivity. However, the sensitivity to color differences is not specific to the categories that correspond to the basic color terms. As a consequence, we conclude that these color categories are not inherent to the basic ability to discriminate colors. Instead, they must develop at a higher, more cognitive level of color processing. At the same time, these results also show that the linguistic distinction between color categories does not affect sensitivity at the category borders through perceptual learning.

Finally, classical studies on categorical perception of color focused on particular stimulus sets and controlled insufficiently for differences in sensitivity. Our results show that under certain conditions some of the categories roughly coincide with the strong variations of JNDs across the hue circle. This is particularly true for the green-blue boundary that has been used as the prime example in many previous studies (cf. Figure 9; see also Witzel & Gegenfurtner, 2011). The coincidence of the categories with patterns of sensitivity may explain why these studies could consistently reproduce patterns that look like category effects (see also A. M. Brown et al., 2011; Witzel & Gegenfurtner, 2011; Lindsey et al., 2010). With respect to these classical studies, our findings point to the importance of controlling low-level perceptual differences, and examining the full set of categories in order

to avoid spurious category effects at higher levels of processing.

Keywords: *categorical perception, color categorization, color discrimination, color naming, language and perception, Sapir-Whorf-hypothesis, second-stage mechanisms*

Acknowledgments

We thank Walter Kirchner for technical assistance, Claudia Kubicek for assistance with data collection, and Thorsten Hansen, Anna Franklin, John Maule, and Lewis Forder for helpful discussion. This research was supported by the German Science Foundation (DFG) Reinhart-Koselleck program Ge 879/9, by a Gießen University dissertation fellowship to C.W., and by the DFG Graduiertenkolleg GRK 885 “NeuroAct.”

Commercial relationships: none.

Corresponding author: Christoph Witzel.

Email: christoph.witzel@psychol.uni-giessen.de.

Address: Department of Psychology, University of Giessen, Giessen, Germany.

References

- Bachy, R., Dias, J., Alleysson, D., & Bonnardel, V. (2012). Hue discrimination, unique hues and naming. *Journal of the Optical Society of America, A: Optics, Image Science, & Vision*, 29(2), A60–A68, doi:10.1364/JOSAA.29.000A60.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley, CA: University of California Press.
- Bornstein, M. H. (1976). Name codes and color memory. *American Journal of Psychology*, 89(2), 269–279.
- Bornstein, M. H., Kessen, W., & Weiskopf, S. (1976). The categories of hue in infancy. *Science*, 191(4223), 201–202.
- Bornstein, M. H., & Korda, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological Research*, 46(3), 207–222.
- Bornstein, M. H., & Monroe, M. D. (1980). Chromatic information processing: Rate depends on stimulus location in the category and psychological complexity. *Psychological Research*, 42(3), 213–225.
- Boynton, R. M., Fargo, L., Olson, C. X., & Smallman,

- H. S. (1989). Category effects in color memory. *Color Research & Application*, 14(5), 229–234.
- Boynton, R. M., & Kambe, N. (1980). Chromatic difference steps of moderate size measured along theoretically critical axes. *Color Research & Application*, 5(1), 13–23.
- Boynton, R. M., & Olson, C. X. (1990). Saliency of chromatic basic color terms confirmed by three measures. *Vision Research*, 30(9), 1311–1317, doi:0042-6989(90)90005-6 [pii].
- Brainard, D. H. (1996). Cone contrast and opponent modulation color spaces. In P. K. Kaiser & R. M. Boynton (Eds.), *Human color vision* (2nd ed., pp. 563–579). Washington, DC: Optical Society of America.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.
- Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *Journal of Neuroscience*, 29(44), 13992–14003, doi:10.1523/JNEUROSCI.3577-09.2009.
- Brown, A. M., Lindsey, D. T., & Guckes, K. M. (2011). Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical. *Journal of Vision*, 11(12):2, 1–21, <http://www.journalofvision.org/content/11/12/2>, doi:10.1167/11.12.2. [PubMed] [Article]
- Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *Journal of Abnormal & Social Psychology*, 49(3), 454–462.
- Cao, D., Pokorny, J., & Smith, V. C. (2005). Associating color appearance with the cone chromaticity space. *Vision Research*, 45(15), 1929–1934, doi:S0042-6989(05)00065-9 [pii] 10.1016/j.visres.2005.01.033.
- Chatterjee, S., & Callaway, E. M. (2003). Parallel colour-opponent pathways to primary visual cortex. *Nature*, 426(6967), 668–671, doi:10.1038/nature02167.
- Crawford, T. D. (1982). Defining “basic color term.” *Anthropological Linguistics*, 24, 338–343.
- Dacey, D. M. (2000). Parallel pathways for spectral coding in primate retina. *Annual Review of Neuroscience*, 23, 743–775.
- Danilova, M. V., & Mollon, J. D. (2010). Parafoveal color discrimination: A chromaticity locus of enhanced discrimination. *Journal of Vision*, 10(1):4, 1–9, <http://www.journalofvision.org/content/10/1/4>, doi:10.1167/10.1.4. [PubMed] [Article]
- Danilova, M. V., & Mollon, J. D. (2012). Foveal color perception: Minimal thresholds at a boundary between perceptual categories. *Vision Research*, 62, 162–172, doi:10.1016/j.visres.2012.04.006.
- Daoutis, C., Pilling, M., & Davies, I. (2006). Categorical effects in visual search for colour. *Visual Cognition*, 14, 217–240.
- De Valois, R. L., Abramov, I., & Jacobs, G. H. (1966). Analysis of response patterns of LGN cells. *Journal of the Optical Society of America A*, 56(7), 966.
- De Valois, R. L., & De Valois, K. K. (1993). A multi-stage color model. *Vision Research*, 33(8), 1053–1065, doi:0042-6989(93)90240-W [pii].
- Demb, J. B., & Brainard, D. H. (2010). Vision: Neurons show their true colours. *Nature*, 467(7316), 670–671, doi:467670b [pii] 10.1038/467670b.
- Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in the lateral geniculate nucleus of macaque. *Journal of Physiology*, 357, 241–265.
- Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., et al. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences, USA*, 104(3), 1097–1102, doi:0610132104 [pii] 10.1073/pnas.0610132104.
- Eskew, R. T., Jr. (2009). Higher order color mechanisms: A critical review. *Vision Research*, 49(22), 2686–2704, doi:S0042-6989(09)00328-9 [pii] 10.1016/j.visres.2009.07.005.
- Fairchild, M. D. (1998). *Color appearance models*. Reading, MA: Addison-Wesley.
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage.
- Field, G. D., Gauthier, J. L., Sher, A., Greschner, M., Machado, T. A., Jepson, L. H., et al. (2010). Functional connectivity in the retina at the resolution of photoreceptors. *Nature*, 467(7316), 673–677, doi:nature09424 [pii] 10.1038/nature09424.
- Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4), 507–521.
- Fisher, R. A. (1921). On the ‘probable error’ of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3–32.
- Fitzgibbon, A. W., Pilu, M., & Fisher, R. B. (1999). Direct least-squares fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5), 476–480.
- Fonteneau, E., & Davidoff, J. (2007). Neural correlates

- of colour categories. *Neuroreport*, 18(13), 1323–1327, doi:10.1097/WNR.0b013e3282c48c33 00001756-200708270-00005 [pii].
- Franklin, A., & Davies, I. R. L. (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology*, 22, 349–377.
- Franklin, A., Drivonikou, G. V., Bevis, L., Davies, I. R. L., Kay, P., & Regier, T. (2008). Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National Academy of Sciences, USA*, 105(9), 3221–3225, doi:0712286105 [pii] 10.1073/pnas.0712286105.
- Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. L. (2008). Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Sciences, USA*, 105(47), 18221–18225, doi:0809952105 [pii] 10.1073/pnas.0809952105.
- Franklin, A., Pilling, M., & Davies, I. R. L. (2005). The nature of infant color categorization: Evidence from eye movements on a target detection task. *Journal of Experimental Child Psychology*, 91(3), 227–248, doi:S0022-0965(05)00053-6 [pii] 10.1016/j.jecp.2005.03.003.
- Gegenfurtner, K. R. (2003). Cortical mechanisms of colour vision. *Nature Reviews Neuroscience*, 4, 563–572.
- Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annual Review of Neuroscience*, 26(1), 181–206.
- Gentner, D., & Goldin-Meadow, S. (2003). Whiter Whorf. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 3–14). Cambridge, MA: MIT Press.
- Giesel, M., Hansen, T., & Gegenfurtner, K. R. (2009). The discrimination of chromatic textures. *Journal of Vision*, 9(9):11, 1–28, <http://www.journalofvision.org/content/9/9/11>, doi:10.1167/9.9.11. [PubMed] [Article]
- Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not in the left. *Proceedings of the National Academy of Sciences, USA*, 103(2), 489–494, doi: 0509868103 [pii] 10.1073/pnas.0509868103.
- Guest, S., & Van Laar, D. (2000). The structure of colour naming space. *Vision Research*, 40(7), 723–734, doi:S0042-6989(99)00221-7 [pii].
- Hansen, T., & Gegenfurtner, K. R. (2013). Higher order color mechanisms: Evidence from noise-masking experiments in cone contrast space. *Journal of Vision*, 13(1):26, 1–21, <http://www.journalofvision.org/content/13/1/26>, doi:10.1167/13.1.26. [PubMed] [Article]
- Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on color categories and color constancy. *Journal of Vision*, 7(4):2, 1–15. <http://www.journalofvision.org/content/7/4/2>, doi:10.1167/7.4.2. [PubMed] [Article]
- Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition*. New York: Cambridge University Press.
- Holmes, A., Franklin, A., Clifford, A., & Davies, I. R. L. (2009). Neurophysiological evidence for categorical perception of color. *Brain & Cognition*, 69(2), 426–434, doi:S0278-2626(08)00288-1 [pii] 10.1016/j.bandc.2008.09.003.
- Hunt, R. W. G. (1980). A model of colour vision for predicting colour appearance. *Color Research & Application*, 7, 95–112.
- Hunt, R. W. G., & Pointer, M. R. (2011). *Measuring colour* (4th ed.). Chichester, UK: John Wiley & Sons.
- IBM Corp. (Released 2011). IBM SPSS Statistics for Windows (Version 20.0). Armonk, NY: IBM Corp.
- Ishihara, S. (2004). *Ishihara's tests for colour deficiency*. Tokyo, Japan: Kanehara Trading Inc.
- Jäkel, F., & Wichmann, F. A. (2006). Spatial four-alternative forced-choice method is the preferred psychophysical method for naive observers. *Journal of Vision*, 6(11):13, 1307–1322, <http://www.journalofvision.org/content/6/11/13>, doi:10.1167/6.11.13. [PubMed] [Article]
- Jameson, K. A., & Komarova, N. L. (2009a). Evolutionary models of color categorization. I. Population categorization systems based on normal and dichromat observers. *Journal of the Optical Society of America A*, 26(6), 1414–1423. doi:180032 [pii].
- Jameson, K. A., & Komarova, N. L. (2009b). Evolutionary models of color categorization. II. Realistic observer models and population heterogeneity. *Journal of the Optical Society of America A*, 26(6), 1424–1436, doi: 180033 [pii].
- Judd, D. B. (1951). *Report of U. S. Secretariat Committee on colorimetry and artificial daylight* (pp. 11). Paris: Bureau Central de la CIE.
- Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis. *American Anthropologist*, 86, 65–79.
- Kay, P., & McDaniel, C. K. (1978). The linguistic

- significance of the meanings of basic color terms. *Language*, 54(3), 610–646.
- Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences*, 100(15), 9085–9089.
- Kay, P., & Regier, T. (2006). Language, thought and color: Recent developments. *Trends in Cognitive Sciences*, 10(2), 51–54, doi:S1364-6613(05)00353-0 [pii] 10.1016/j.tics.2005.12.007.
- Knight, R., Buck, S. L., Fowler, G. A., & Nguyen, A. (1998). Rods affect S-cone discrimination on the Farnsworth-Munsell 100-hue test. *Vision Research*, 38(21), 3477.
- Komarova, N. L., & Jameson, K. A. (2008). Population heterogeneity and color stimulus heterogeneity in agent-based color categorization. *Journal of Theoretical Biology*, 253(4), 680–700, doi:S0022-5193(08)00154-9 [pii] 10.1016/j.jtbi.2008.03.030.
- Komarova, N. L., Jameson, K. A., & Narens, L. (2007). Evolutionary models of color categorization based on discrimination. *Journal of Mathematical Psychology*, 51(6), 359–382.
- Krauskopf, J. (1999). Higher order color mechanism. In K. R. Gegenfurtner & L. T. Sharpe (Eds.), *Color vision: From genes to perception* (pp. 303–316). Cambridge, UK: Cambridge University Press.
- Krauskopf, J., & Gegenfurtner, K. R. (1992). Color discrimination and adaptation. *Vision Research*, 32(11), 2165–2175.
- Krauskopf, J., Williams, D. R., & Heeley, D. W. (1982). Cardinal directions of color space. *Vision Research*, 22(9), 1123–1131.
- Kuehni, R. G., & Schwarz, A. (2008). *Color ordered: A survey of color order systems from antiquity to the present*. New York: Oxford University Press.
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93–107.
- Laws, G., Davies, I., & Andrews, C. (1995). Linguistic structure and non-linguistic cognition: English and Russian blues compared. *Language & Cognitive Processes*, 10(1), 59–94.
- Lee, B. B., Martin, P. R., & Valberg, A. (1988). The physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina. *Journal of Physiology*, 404, 323.
- Lennie, P., Krauskopf, J., & Sclar, G. (1990). Chromatic mechanisms in striate cortex of macaque. *Journal of Neuroscience*, 10(2), 649–669.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49(2), 467–477.
- Lindsey, D. T., Brown, A. M., Reijnen, E., Rich, A. N., Kuzmova, Y. I., & Wolfe, J. M. (2010). Color channels, not color appearance or color categories, guide visual search for desaturated color targets. *Psychological Science*, 21(9), 1208–1214, doi: 0956797610379861 [pii] 10.1177/0956797610379861.
- Linhares, J. M., Pinto, P. D., & Nascimento, S. M. (2008). The number of discernible colors in natural scenes. *Journal of the Optical Society of America A*, 25(12), 2918–2924, doi:173260 [pii].
- MacLeod, D. I. A., & Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America*, 69(8), 1183–1186.
- Malkoc, G., Kay, P., & Webster, M. A. (2005). Variations in normal color vision. IV. Binary hues and hue scaling. *Journal of the Optical Society of America A*, 22(10), 2154–2168.
- Munsell Color Services. (2007a). *The Munsell book of color: Glossy collection*. Grandville, MI: x-rite.
- Munsell Color Services. (2007b). *The Munsell book of color: Matte collection*. Grandville, MI: x-rite.
- Neitz, J., & Neitz, M. (2011). The genetics of normal and defective color vision. *Vision Research*, doi: S0042-6989(10)00569-9 [pii] 10.1016/j.visres.2010.12.002.
- Neitz, M., Neitz, J., & Jacobs, G. H. (1991). Spectral tuning of pigments underlying red-green color vision. *Science*, 252(5008), 971–974.
- Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2010). Categorical color constancy for real surfaces. *Journal of Vision*, 10(9):16, 1–22, <http://www.journalofvision.org/content/10/9/16>, doi:10.1167/10.9.16. [PubMed] [Article]
- Özgen, E. (2004). Language, learning, and color perception. *Current Directions in Psychological Science*, 13(3), 95–98.
- Özgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, 131(4), 477–493.
- Parkes, L. M., Marsman, J. B., Oxley, D. C., Goulermas, J. Y., & Wuerger, S. M. (2009). Multivoxel fMRI analysis of color tuning in human primary visual cortex. *Journal of Vision*, 9(1):1, 1–13, <http://www.journalofvision.org/content/9/1/1>, doi:10.1167/9.1.1. [PubMed] [Article]
- Pelli, D. G. (1997). The VideoToolbox software for

- visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Pilling, M., Wiggett, A., Özgen, E., & Davies, I. R. L. (2003). Is color “categorical perception” really perceptual? *Memory & Cognition*, 31(4), 538–551.
- Pinto, L., Kay, P., & Webster, M. A. (2010). Color categories and perceptual grouping. *Journal of Vision*, 10(7):409, <http://www.journalofvision.org/content/10/7/409>, doi:10.1167/10.7.409. [Abstract]
- Pointer, M. R., & Attridge, G. G. (1998). The number of discernible colours. *Color Research & Application*, 23(1), 52–54.
- Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences, USA*, 104(4), 1436–1441, doi:0610341104 [pii] 10.1073/pnas.0610341104.
- Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, 50(4), 378–411. doi:S0010-0285(04)00076-3 [pii] 10.1016/j.cogpsych.2004.10.001.
- Roberson, D., Davies, I. R. L., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, 129(3), 369–398.
- Roberson, D., & Hanley, J. R. (2007). Color vision: Color categories vary with language after all. *Current Biology*, 17(15), R605–R607, doi:S0960-9822(07)01481-9 [pii] 10.1016/j.cub.2007.05.057.
- Roberson, D., Hanley, J. R., & Pak, H. (2009). Thresholds for color discrimination in English and Korean speakers. *Cognition*, 112(3), 482–487, doi: S0010-0277(09)00139-5 [pii] 10.1016/j.cognition.2009.06.008.
- Sapir, E. (1921). *Language: An introduction to the study of speech*. New York: Harcourt.
- Schnapf, J. L., Kraft, T. W., & Baylor, D. A. (1987). Spectral sensitivity of human cone photoreceptors. *Nature*, 325(6103), 439–441, doi:10.1038/325439a0.
- Shinoda, H., Uchikawa, K., & Ikeda, M. (1993). Categorized color space on CRT in the aperture and the surface color mode. *Color Research & Application*, 18(5), 326–333.
- Siok, W. T., Kay, P., Wang, W. S. Y., Chan, A. H. D., Chen, L., Luke, K.-K., et al. (2009). Language regions of brain are operative in color perception. *Proceedings of the National Academy of Sciences, USA*, 106(20), 8140–8145, doi:0903627106 [pii] 10.1073/pnas.0903627106.
- Smith, V. C., & Pokorny, J. (1975). Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision Research*, 15(2), 161.
- Stabell, B., & Stabell, U. (1998). Chromatic rod-cone interaction during dark adaptation. *Journal of the Optical Society of America A*, 15(11), 2809.
- Stabell, B., & Stabell, U. (2002). Effects of rod activity on color perception with light adaptation. *Journal of the Optical Society of America A*, 19(7), 1249.
- Stockman, A., & Brainard, D. H. (2010). Color vision mechanisms. In M. Bass (Ed.), *OSA Handbook of Optics* (3rd ed., pp. 11.11–11.104). New York: McGraw-Hill.
- Stockman, A., & Sharpe, L. T. (2000). The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research*, 40(13), 1711–1737, doi:S0042-6989(00)00021-3 [pii].
- Sturges, J., & Whitfield, T. W. A. (1997). Salient features of Munsell colour space as a function of monolexic naming and response latencies. *Vision Research*, 37(3), 307–313, doi:S0042-6989(96)00170-8 [pii].
- Tailby, C., Solomon, S. G., & Lennie, P. (2008). *Functional asymmetries in visual pathways carrying S-cone signals in macaque*. *Journal of Neuroscience*, 28(15), 4078–4087, doi:10.1523/JNEUROSCI.5338-07.2008.
- The MathWorks Inc. (2007). *Matlab: The language of technical computing* (Version R2007a). Natick, MA: The MathWorks Inc.
- Trujillo-Ortiz, A., Hernandez-Walls, R., & Trujillo-Perez, R. A. (Producers). (2004). RMAOV2: Two-way repeated measures ANOVA. Retrieved from <http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=5578>.
- Uchikawa, H., Uchikawa, K., & Boynton, R. M. (1989). Influence of achromatic surrounds on categorical perception of surface colors. *Vision Research*, 29(7), 881–890.
- Uchikawa, K., & Shinoda, H. (1996). Influence of basic color categories on color memory discrimination. *Color Research & Application*, 21(6), 430–439.
- Valberg, A. (2001). Unique hues: An old problem for a new generation. *Vision Research*, 41(13), 1645–1657, doi:S0042-6989(01)00041-4 [pii].
- Webster, M. A., Miyahara, E., Malkoc, G., & Raker, V. E. (2000). Variations in normal color vision. I. Cone-opponent axes. *Journal of the Optical Society of America A*, 17(9), 1535–1544.
- Whorf, B. L. (1964). *Language, thought and reality*. Cambridge, MA: MIT Press.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L.,

- Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, USA*, 104(19), 7780–7785, doi:0701644104 [pii] 10.1073/pnas.0701644104.
- Witthoft, N., Winawer, J., Wu, L., Frank, M., Wade, A., & Boroditsky, L. (2003). *Effects of language on color discriminability*. Paper presented at the 25th Annual Meeting of the Cognitive Science Society, Mahwah, NJ.
- Witzel, C., & Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *Journal of Vision*, 11(12):16, 1–25, <http://www.journalofvision.org/content/11/12/16>, doi:10.1167/11.12.16. [PubMed] [Article]
- Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2008a). Categorical discrimination of colour. *Journal of Vision*, 8(6): 577, <http://www.journalofvision.org/content/8/6/577>, doi:10.1167/8.6.577. [Abstract]
- Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2008b). *Wie sich Farben mit den Betrachtern und mit den Zeiten ändern* [Translation: *How colors change across observers and time*]. Paper presented at the Tagung experimentell arbeitender Psychologen (TeaP), Marburg, Germany.
- Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae* (2nd ed.). New York: John Wiley & Sons.
- Yokoi, K., Nishimori, T., & Saida, S. (2008). Interference of verbal labels in color categorical perception. *Optical Review*, 15(6), 295–301.
- Yokoi, K., & Uchikawa, K. (2005). Color category influences heterogeneous visual search for color. *Journal of the Optical Society of America A*, 22(11), 2309–2317.
- Zeile, A. J., Kremers, J., & Feigl, B. (2012). Mesopic rod and S-cone interactions revealed by modulation thresholds. *Journal of the Optical Society of America A: Optics, Image Science, & Vision*, 29(2), A19–26, doi:10.1364/JOSAA.29.000A19.